

# Optimierungsmöglichkeiten beim Selbstarchivieren wissenschaftlicher Publikationen in Institutionellen Repositorien

Masterarbeit

Studiengang Bibliotheks- und Informationswissenschaft

Master in Library and Information Science

Fakultät für Informations- und Kommunikationswissenschaften  
der Fachhochschule Köln

vorgelegt von:

Stefan Grund (Matrikelnr.: 11069460)

am 11. August 2011 bei Dr. Heike Neuroth

Zweitprüfer: Prof. Dr. Achim Oßwald

Hiermit versichere ich, die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt zu haben.

Frankfurt am Main, den 11. August 2011

(Stefan Grund)

## Abstract (dt.)

Um eine möglichst große Zahl von wissenschaftlichen Publikationen für jeden frei zugänglich im Internet zur Verfügung zu stellen, verfolgt die Open-Access-Bewegung neben der Gründung von freien Online-Zeitschriften auch die Strategie der Selbstarchivierung: Wissenschaftler sollen ihre in wissenschaftlichen Peer-Review-Zeitschriften veröffentlichten Publikationen – unter der Einhaltung urheberrechtlicher Bestimmungen – in digitalen Repositorien zur Verfügung stellen, die mittlerweile von vielen Universitäten und anderen Forschungsinstitutionen eingerichtet wurden. Aktuelle Untersuchungen zeigen allerdings, dass nur wenige Wissenschaftler dieser Aufforderung Folge leisten. Um Wissenschaftler zu motivieren, die Ablieferungsrate von Veröffentlichungen zu erhöhen, ist es daher wichtig, dass Repositorien einen Mehrwert für den Wissenschaftsbetrieb zur Verfügung stellen, und dass der Prozess der Erfassung von Metadaten in Repositorien vereinfacht wird. Die vorliegende Arbeit untersucht, welche Qualitätskriterien Metadaten erfüllen müssen, um Repositorien durch umfangreichere Dienstleistungen und eine bessere Vernetzung zu einem effizienten Hilfsmittel für die Wissenschaft zu machen. Dabei greifen wir zuerst zurück auf bereits veröffentlichte Befunde aus quantitativen und qualitativen Umfragen, die unter Wissenschaftlern durchgeführt wurden, und führen dann eine Analyse der Veränderungen in der modernen Wissenschaftskommunikation durch, um vorstellbare und notwendige Dienstleistungen zu diskutieren. Nach einer Darstellung allgemeiner Qualitätskriterien für Metadaten analysieren wir, in welcher Art solche Kriterien in Repositorien angewendet werden können. Zudem werden funktionale Anforderungen für die einzelnen Elemente auf der Grundlage der vorher entworfenen Services erstellt. Zuletzt werden Hilfsmittel beschrieben, die eingesetzt werden sollten, um den Aufwand für die Wissenschaftler bei der Erfassung der Metadaten für ihre Publikationen soweit wie möglich zu reduzieren.

**Schlagwörter:** Open Access, Selbstarchivierung, Institutionelle Repositorien, Metadaten, Wissenschaftskommunikation

## Abstract (eng.)

The open access movement seeks to make as many scientific publications as possible freely available on the internet. In order to achieve this goal, besides establishing online journals which are free of charge, it follows the strategy of self-archiving. According to this strategy, scientists are required to make their peer-reviewed journal publications available in digital repositories while complying with copyright regulations. Meanwhile,

such repositories have been established by many universities and research institutes. However, recent studies have shown that most academics do not make their publications available. In order to motivate scientists to deliver their publications to repositories and to record metadata, it is important that repositories provide an added value for scientists and that recording metadata about their studies becomes easier. The current study explores which quality criteria metadata must fulfill so that repositories can become an efficient tool for scientists and academics. In addition, it is argued that repositories must provide more extensive services and better interconnectedness to achieve this goal. I first discuss published findings from quantitative and qualitative surveys, which were conducted among scientists, to propose conceivable service models. After presenting general quality criteria for metadata, I illustrate how such criteria may be applied in repositories. Furthermore, functional requirements for the single elements of metadata are outlined. Finally, I describe procedures that should be introduced to decrease the efforts required by scientists as far as possible in recording metadata of their publications.

**Keywords:** Open Access. Self-archiving, Institutional repository, Metadata, Scholarly communication

# Inhaltsverzeichnis

|  |             |
|--|-------------|
| <b>Abbildungsverzeichnis</b>   | <b>VIII</b> |
| <b>Tabellenverzeichnis</b>   | <b>VIII</b> |
| <b>Abkürzungsverzeichnis</b>   | <b>VIII</b> |
| <b>Startseiten behandelter Repositorien</b>  | <b>X</b>    |
| <b>1 Einleitung</b>  | <b>1</b>    |
| <b>2 Open Access und neue Formen der Wissenschaftskommunikation</b>  | <b>4</b>    |
| 2.1 Der Open-Access-Gedanke . . . . .  | 4           |
| 2.1.1 Der goldene Weg des Open Access: Alternative Publikationsformen  | 6           |
| 2.1.2 Der grüne Weg des Open Access: Selbstarchivierung . . . . .  | 7           |
| 2.2 Anforderungen an Institutionelle Repositorien aus Sicht der Wissenschaft   | 10          |
| 2.2.1 Die Einstellung von Wissenschaftlern zu Open Access und Selbstarchivierung . . . . .   | 10          |
| 2.2.2 Individuelle Mehrwertdienste für Wissenschaftler . . . . .   | 15          |
| 2.2.3 Kollaborative Mehrwertdienste im Rahmen neuer Formen der Wissenschaftskommunikation . . . . .  | 17          |
| 2.2.4 Die „Thesen zur Zukunft von Open-Access-Repositorien“ . . . . .  | 20          |
| <b>3 Die Erschließung selbstarchivierter Dokumente in Institutionellen Repositorien</b>  | <b>22</b>   |
| 3.1 Bedeutung von Metadaten für Elektronische Dokumente . . . . .  | 22          |
| 3.2 Derzeitige Situation der Metadatenerschließung in Institutionellen Repositorien . . . . .  | 24          |
| <b>4 Qualität von Metadaten</b>  | <b>28</b>   |
| 4.1 Definition . . . . .   | 28          |
| 4.2 Allgemeine Qualitätskriterien für Metadaten elektronischer Ressourcen .  | 30          |
| 4.3 Allgemeine Qualitätskriterien für Metadaten elektronischer Dokumente nach Bruce/Hillmann . . . . .   | 32          |
| 4.4 Funktionale Anforderungen . . . . .  | 35          |
| <b>5 Qualitätskriterien und funktionale Anforderungen an Metadaten für selbstarchivierte Zeitschriftenpublikationen in Institutionellen Repositorien</b> | <b>38</b>   |

|       |   |    |
|-------|---|----|
| 5.1   | Elementübergreifende Kriterien . . . . .              | 39 |
| 5.1.1 | Vorgaben DINI/DRIVER . . . . .                        | 39 |
| 5.1.2 | Allgemeine Qualitätskriterien . . . . .               | 39 |
| 5.1.3 | Funktionale Anforderungen . . . . .                   | 40 |
| 5.2   | Titel . . . . .                                       | 40 |
| 5.2.1 | Vorgaben DC Simple/SWAP . . . . .                     | 41 |
| 5.2.2 | Vorgaben DINI/DRIVER . . . . .                        | 41 |
| 5.2.3 | Allgemeine Qualitätskriterien . . . . .               | 41 |
| 5.2.4 | Funktionale Anforderungen . . . . .                   | 41 |
| 5.3   | Verfasser . . . . .                                   | 41 |
| 5.3.1 | Vorgaben DC Simple/SWAP . . . . .                     | 42 |
| 5.3.2 | Vorgaben DINI/DRIVER . . . . .                        | 42 |
| 5.3.3 | Allgemeine Qualitätskriterien . . . . .               | 42 |
| 5.3.4 | Funktionale Anforderungen . . . . .                   | 43 |
| 5.4   | Sachliche/Inhaltliche Erschließung . . . . .          | 47 |
| 5.4.1 | Vorgaben DC Simple/SWAP . . . . .                     | 47 |
| 5.4.2 | Vorgaben DINI/DRIVER . . . . .                        | 48 |
| 5.4.3 | Allgemeine Qualitätskriterien . . . . .               | 48 |
| 5.4.4 | Funktionale Anforderungen . . . . .                   | 49 |
| 5.5   | Quelle . . . . .                                      | 51 |
| 5.5.1 | Vorgaben DC Simple/SWAP . . . . .                     | 51 |
| 5.5.2 | Vorgaben DINI/DRIVER . . . . .                        | 52 |
| 5.5.3 | Allgemeine Qualitätskriterien . . . . .               | 52 |
| 5.5.4 | Funktionale Anforderungen . . . . .                   | 53 |
| 5.6   | Affiliation . . . . .                                 | 54 |
| 5.6.1 | Vorgaben DC Simple/SWAP . . . . .                     | 54 |
| 5.6.2 | Vorgaben DINI/DRIVER . . . . .                        | 55 |
| 5.6.3 | Allgemeine Qualitätskriterien . . . . .               | 55 |
| 5.6.4 | Funktionale Anforderungen . . . . .                   | 55 |
| 5.7   | Eigenschaften der hochgeladenen Publikation . . . . . | 56 |
| 5.7.1 | Vorgaben DC Simple/SWAP . . . . .                     | 56 |
| 5.7.2 | Vorgaben DINI/DRIVER . . . . .                        | 57 |
| 5.7.3 | Allgemeine Qualitätskriterien . . . . .               | 58 |
| 5.7.4 | Funktionale Anforderungen . . . . .                   | 58 |
| 5.8   | Fazit . . . . .                                       | 60 |

|          |   |           |
|----------|---|-----------|
| <b>6</b> | <b>Unterstützungsmöglichkeiten zur Verbesserung der Erschließung selbstarchivierter Dokumente</b> | <b>62</b> |
| 6.1      | vor Submitvorgang . . . . .   | 63        |
| 6.2      | während Submitvorgang . . . . .   | 64        |
| 6.2.1    | Nutzerprofil . . . . .  | 64        |
| 6.2.2    | Import von Datensätzen . . . . .  | 64        |
| 6.2.3    | Maschinell aus dem hochgeladenen Dokument erzeugte Metadaten                                      | 65        |
| 6.2.4    | Nutzung von APIs zur Verknüpfung mit Normdaten . . . . .  | 66        |
| 6.2.5    | Vorschlaglisten in den Eingabefeldern . . . . .   | 67        |
| 6.2.6    | Validierung der Eingaben . . . . .  | 68        |
| 6.3      | nach Submitvorgang . . . . .  | 68        |
| 6.3.1    | Manuelle Endredaktion . . . . .   | 68        |
| 6.3.2    | Automatisierte Anreicherungen . . . . .   | 68        |
| 6.4      | Fazit . . . . .   | 69        |
| <b>7</b> | <b>Schlussbetrachtung</b>   | <b>74</b> |
|          | <b>Literatur</b>  | <b>78</b> |
| <b>A</b> | <b>Verwendete Datensätze</b>  | <b>94</b> |
| A.1      | Datensatz <i>Goescholar</i> . . . . .   | 94        |
| A.2      | Datensatz <i>BiPrints</i> . . . . .   | 97        |
| A.3      | Datensatz <i>KOPS</i> . . . . .   | 100       |
| A.4      | Datensatz <i>Edocs</i> HU . . . . .   | 103       |
| A.5      | Datensatz <i>ZORA</i> . . . . .   | 105       |
| A.6      | Datensatz <i>CrossRef</i> . . . . .   | 107       |
| A.7      | Datensatz <i>PND</i> . . . . .  | 109       |
| A.8      | Datensatz <i>LCSH</i> . . . . .   | 110       |
| A.9      | Datensatz <i>SWD</i> . . . . .  | 111       |
| A.10     | Datensatz <i>GKD</i> . . . . .  | 112       |

## Abbildungsverzeichnis

|   |  |    |
|---|--|----|
| 1 | <i>Beispiele Eingabeformular</i> . . . . .                           | 25 |
| 2 | <i>Ebenen von Metadatensätzen aus Repositorien</i> . . . . .         | 37 |
| 3 | <i>Zuordnung verschiedener Namensformen über Normdaten</i> . . . . . | 46 |

## Tabellenverzeichnis

|   |  |    |
|---|--|----|
| 1 | <i>Verschiedene Stufen eines Metadatensatzes</i> . . . . . | 71 |
|---|--|----|

## Abkürzungsverzeichnis

|               |  |
|---------------|--|
| <b>AACR2</b>  | Anglo-American Cataloging Rules, 2. ed.                        |
| <b>BOAI</b>   | Budapest Open Access Initiative                                |
| <b>BSZ</b>    | Bibliotheksservice-Zentrum Baden-Württemberg                   |
| <b>CTA</b>    | Copyright Transfer Agreement                                   |
| <b>DAI</b>    | Digitale Auteur Identifier                                     |
| <b>DC</b>     | Dublin Core  |
| <b>DDC</b>    | Dewey Decimal Classification                                   |
| <b>DFG</b>    | Deutsche Forschungsgemeinschaft                                |
| <b>DINI</b>   | Deutsche Initiative für Netzwerkinformation                    |
| <b>DNB</b>    | Deutsche Nationalbibliothek                                    |
| <b>DOI</b>    | Digital Object Identifier                                      |
| <b>DRIVER</b> | Digital Repository Infrastructure Vision for European Research |
| <b>EZB</b>    | Elektronische Zeitschriftenbibliothek                          |
| <b>GKD</b>    | Gemeinsame Körperschaftsdatei                                  |
| <b>GND</b>    | Gemeinsame Normdatei   |
| <b>IR</b>     | Institutionelle Repositorien                                   |



|                |   |
|----------------|---|
| <b>ISNI</b>    | International Standard Name Identifier  |
| <b>KOPS</b>    | Konstanzer Online-Publikations-System   |
| <b>IFLA</b>    | International Federation of Library Associations and Institutions               |
| <b>JISC</b>    | Joint Information Systems Committee   |
| <b>LCSH</b>    | Library of Congress Subject Headings  |
| <b>LoC</b>     | Library of Congress   |
| <b>MACS</b>    | Multilingual Access to Subjects   |
| <b>METS</b>    | Metadata Encoding and Transmission Standard                                     |
| <b>MODS</b>    | Metadata Object Description Schema  |
| <b>OAI-PMH</b> | Open Archives Initiative Protocol for Metadata Harvesting                       |
| <b>OAI-ORE</b> | Open Archives Initiative Object Reuse and Exchange                              |
| <b>OSWD</b>    | Online-SWD  |
| <b>PND</b>     | Personennamendatei  |
| <b>RAK-WB</b>  | Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken |
| <b>RAMEAU</b>  | Répertoire d'autorité matière encyclopédique et alphabétique unifié             |
| <b>RDF</b>     | Resource Description Framework  |
| <b>RSWK</b>    | Regeln für den Schlagwortkatalog  |
| <b>SRU</b>     | Search/Retrieve via URL   |
| <b>SSOAR</b>   | Social Science Open Access Repository   |
| <b>SWAP</b>    | Scholarly Works Application Profile   |
| <b>STW</b>     | Standard Thesaurus Wirtschaft   |
| <b>SWD</b>     | Schlagwortnormdatei   |
| <b>URN</b>     | Uniform Resource Name   |

|             |                                      |
|-------------|--------------------------------------|
| <b>VIAF</b> | Virtual International Authority File |
| <b>ZDB</b>  | Zeitschriftendatenbank               |
| <b>ZORA</b> | Zurich Open Repository and Archive   |

## Startseiten behandelter Repositorien

|                                 |   |
|---------------------------------|---|
| arXiv (Physik, Mathematik etc.) | <a href="http://arxiv.org">http://arxiv.org</a>   |
| BiPrints (Univ. Bielefeld)      | <a href="http://repositories.ub.uni-bielefeld.de/biprints/index.php?la=de">http://repositories.ub.uni-bielefeld.de/biprints/index.php?la=de</a> |
| edoc (HU Berlin)                | <a href="http://edoc.hu-berlin.de/">http://edoc.hu-berlin.de/</a>   |
| FreiDok (Univ. Freiburg)        | <a href="http://www.freidok.uni-freiburg.de/?la=de">http://www.freidok.uni-freiburg.de/?la=de</a>   |
| Goescholar (Univ. Göttingen)    | <a href="http://goedoc.uni-goettingen.de/goescholar">http://goedoc.uni-goettingen.de/goescholar</a>   |
| KOPS (Univ. Konstanz)           | <a href="http://kops.ub.uni-konstanz.de">http://kops.ub.uni-konstanz.de</a>   |
| SSOAR (Sozialwissenschaften)    | <a href="http://www.ssoar.info/">http://www.ssoar.info/</a>   |
| ZORA (Univ. Zürich)             | <a href="http://www.zora.uzh.ch">http://www.zora.uzh.ch</a>   |

# 1 Einleitung

Die in den letzten Jahrzehnten entwickelten technischen Möglichkeiten des Internets haben einen großen Einfluss auf die Veröffentlichungspraxis wissenschaftlicher Forschungsergebnisse. Hat sich die Wissenschaftskommunikation schon in den Jahrzehnten davor immer weiter beschleunigt und ihren Fokus – vor allem in den naturwissenschaftlichen Fächern – von Monographien auf regelmäßig mehrmals im Jahr erscheinende Zeitschriften verlagert<sup>1</sup>, bieten digitale Instrumente noch größere Aktualität und eine zumindest technische Unabhängigkeit vom Verlagswesen und seiner Infrastruktur. Neben dem schnellen, informellen Austausch über Email-Listen oder einschlägige Blogs wurden Online-Zeitschriften gegründet, die weiterhin Peer-Review-Verfahren vorsehen, aber für jeden frei zugänglich sind, ohne dass ein Subskriptionspreis gezahlt werden muss.<sup>2</sup> In bestimmten Fächern, wie z. B. Physik, gewannen ebenfalls öffentlich zugängliche Preprint-Server an Bedeutung, auf denen Publikationen bereits im Entwurfsstatus veröffentlicht und gemeinsam mit der Fachcommunity weiterentwickelt werden.<sup>3</sup>

Dennoch haben die durch Impact Factor und erwartete höhere Reputation und Qualität charakterisierten traditionellen Verlagszeitschriften ihre dominierende Position bis heute behalten – hier erscheinen weiterhin ca. 80% der veröffentlichten Artikel, u. U. zwar elektronisch, aber weiterhin nicht frei zugänglich, nicht als *Open Access*.<sup>4</sup> Eine Forderung der *Budapest Open Access Initiative* sieht daher neben der Einrichtung von freien Online-Zeitschriften auch die Option der „Selbstarchivierung“ vor – Verfasser sollen ihre bereits veröffentlichten Verlagspublikationen, u. U. in einer formal abweichenden, aber inhaltlich identischen Form, auf Dokumentenserver, sogenannte *Repositorien*, hochladen und sie so dem Wissenschaftsprozess auf einfache Art und Weise zur Verfügung stellen.<sup>5</sup>

Neben einer kleineren Zahl *fachlich* orientierter Repositorien wurden zur Unterstützung dieses Ziels vor allem *institutionelle* Repositorien gegründet. Ein institutionelles Repositorium (**IR**) lässt sich definieren als das Angebot beispielsweise einer Universität an ihre Mitarbeiter, von ihnen erstellte wissenschaftliche Publikationen digital offen für

---

<sup>1</sup>Vgl. exemplarisch Kölbl (2011), 185f; Gantert u. Hacker (2008), 72f. Zur seit dem 19. Jh. rasant angestiegenen Zahl der wissenschaftlichen Zeitschriften vgl. ebenfalls Gantert u. Hacker (2008), 184.

<sup>2</sup>Eine der ersten dieser neuen Art von Zeitschriften war das *Electronic Journal of Communication* mit Erscheinungsjahr September 1990 (vgl. Open Access Directory (2010)).

<sup>3</sup>Einer der bekanntesten Preprint-Server ist *arXiv*, s. a. S. 9 dieser Arbeit.

<sup>4</sup>Vgl. Björk u. a. (2010), Abschn. 'Results'.

<sup>5</sup>Vgl. BOAI (2002).

jeden zur Verfügung zu stellen, sie aktiv zu verbreiten und für ihre Langzeitarchivierung zu sorgen.<sup>6</sup>

Bibliotheken und universitäre Rechenzentren starteten mit großem Optimismus seit Anfang der 90er Jahre des 20. Jahrhunderts<sup>7</sup> Repositorien, aber: „The ‘build it and they will come’ proposition has been decisively proven wrong“<sup>8</sup>, d. h. Repositorien werden von den Wissenschaftlern nicht so gut angenommen wie erhofft. In der Forschungsliteratur wird daher gefordert, dass Repositorien durch Mehrwert und zusätzliche Dienstleistungen die tägliche Arbeit eines Wissenschaftlers erleichtern sollten und nicht vielmehr durch den Prozess des Hochladens und Erschließens von Dokumenten belasten. In diesem Falle würden die Chancen steigen, dass Wissenschaftler den üblichen Publikationsabläufen einen weiteren selbstverständlichen Schritt „Selbstarchivierung“ hinzufügen.<sup>9</sup> Die Leistung, die ein Repositoryum erbringen kann, steht und fällt allerdings mit der Qualität der Daten, über die es verfügt – dies bezieht sich sowohl auf die Dokumente als auch auf die sie beschreibenden Metadaten: Durch Peer-Review-Verfahren qualitätsgesicherte Artikel sind im Wissenschaftsprozess relevanter als nicht begutachtete; Metadaten, die Dokumente nur unzureichend beschreiben, erschweren deren Auffinden und damit Präsenz in Suchmaschinen und anderen wissenschaftlichen Forschungsumgebungen.

Diese Arbeit möchte daher untersuchen, welche Qualitätskriterien Metadaten erfüllen müssen, um Repositorien durch umfangreichere Dienstleistungen und eine bessere Vernetzung zu einem effizienten Hilfsmittel für die Wissenschaft zu machen. Dabei ist es wichtig zu berücksichtigen, welche Services Forscher von einer Publikationsplattform erwarten und welche Verbesserungen ihrer täglichen Arbeit mit Publikationen und Forschungsdaten sie sich grundsätzlich vorstellen können, ohne dies vielleicht bisher mit Repositorien in Verbindung gebracht zu haben. Auf der Grundlage einer Analyse dieser Erwartungen sollen über die bereits bekannten allgemeinen Qualitätskriterien für Metadaten elektronischer Dokumente und die technischen Ansprüche der Schnittstellen zum Datenaustausch hinaus auch *funktionale Anforderungen* entworfen werden.

Nach einer kurzen Darstellung des Open-Access-Gedankens werden dazu erst die Ergebnisse zweier aktuellen Umfragen ausgewertet, die Wissenschaftler nach ihren Einstellungen zu den verschiedenen Formen von Open Access befragten. Auf ihrer Grundlage werden individuelle Dienstleistungen entworfen. Eine zweite behandelte Basis der

---

<sup>6</sup>Vgl. Lynch (2003), Abschn. ‘Defining Institutional Repositories’. Eine ausführlichere Definition findet sich auf S. 8 dieser Arbeit.

<sup>7</sup>Vgl. Boosen (2008), 16 u. 21.

<sup>8</sup>Salo (2008), 98.

<sup>9</sup>Vgl. Müller u. a. (2009), Abschn. 2; Salo (2008), 120; Foster u. Gibbons (2005), Abschn. ‘Institutional Repositories and the Adoption Problem’.

Konzeption für Anforderungen an Services von Repositorien bezieht sich auf allgemeine Entwicklungen der Wissenschaftskommunikation: Auch für Modelle wie *virtuelle Forschungsumgebungen* oder *eScience* stellen Repositorien wichtige Elemente dar, da sie hier als miteinander vernetzte Ausgangsportale für den Zugriff auf Publikationen dienen sollen.

Im nächsten Teil stehen die Metadaten aus den Repositorien im Mittelpunkt. Die Erschließungsqualität sowohl formaler als auch inhaltlicher Art entscheidet über den Umfang der Services, der erbracht werden kann. Beispielhaft sollen dabei die gängigen Metadatenelemente für Zeitschriftenartikel analysiert werden. Verschiedene Kriterien sowohl technischer – beispielsweise Ansprüche der Schnittstellen zum Austausch von Daten – wie menschlicher Art – beispielsweise die Bereitschaft des Dokumentablieferers, Zeit und Aufwand für die Metadatenerfassung zu erbringen – müssen dabei berücksichtigt werden. Nach der Darstellung allgemeiner Qualitätskriterien für Metadaten elektronischer Ressourcen sowie der bereits existierenden und angewandten Anforderungen durch das *DINI-Zertifikat*<sup>10</sup> und die *DRIVER Guidelines*<sup>11</sup> werden funktionale Anforderungen für die einzelnen Elemente entwickelt, deren Erfüllung für die angedachten Services notwendig ist.

Das Ziel, eine höhere Qualität von Metadaten zu erreichen, führt sehr wahrscheinlich zu einem höheren Aufwand für den Erschließer. Da diese eher als Verwaltungstätigkeit zu charakterisierende Aufgabe einem Wissenschaftler nicht vermittelbar und der Nutzung von Repositorien daher hinderlich ist, müssen Instrumente und Hilfsmittel entwickelt werden, die den Datenerfasser unterstützen und auch nach dem Erschließungsvorgang die Metadatenansätze anreichern und weiter aktuell halten. Hierfür werden im letzten Kapitel Vorschläge entwickelt, die sich auf Usability und Funktionalitäten der Erschließungsmasken sowie eine automatische Verknüpfung mit externen Normdaten zur Standardisierung der Metadatenansätze beziehen.

---

<sup>10</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011).

<sup>11</sup>Vanderfeesten u. a. (2008).

## 2 Open Access und neue Formen der Wissenschaftskommunikation

### 2.1 Der Open-Access-Gedanke

Durch die Entwicklung moderner und allgemein verfügbarer Informations- und Kommunikationstechnologien sind Wissenschaftler und ihre Institutionen mittlerweile in der Lage, unabhängig von der Infrastruktur der Verlage<sup>12</sup> ihre Forschungsergebnisse für jeden frei zugänglich, also *Open Access*, zu publizieren. Der klassische Prozess von öffentlich geförderter Forschung, aber kommerzieller Veröffentlichung ihrer Ergebnisse, kann so gebrochen werden. Besonders Forschungsinstitutionen haben ein großes Interesse daran, die Publikation von Forschungsergebnissen ihrer Mitarbeiter und Mitarbeiterinnen nicht für teuer Geld zurückkaufen zu müssen, da die seit den letzten Jahrzehnten des 20. Jahrhunderts rasant gestiegenen Zeitschriftenpreise die Budgets der Bibliotheken stark belasten und zur Abbestellung von Subskriptionen führen. Die sinkende Anzahl von Abonnements hat weitere Preiserhöhungen und damit wiederum Abbestellungen zur Folge – eine kritische Situation, die allgemein als „Zeitschriftenkrise“<sup>13</sup> bekannt ist. So sind beispielsweise die Preise der sozialwissenschaftlichen Zeitschriften wichtiger Verlage im Zeitraum von 2000 – 2006 um 47% (Elsevier) – 120% (Univ. of Chicago Press) gestiegen.<sup>14</sup>

Der Großteil der relevanten Zeitschriften befindet sich mittlerweile in der Hand weniger Verlage, die versuchen, ihre Bedingungen am Markt durchzusetzen (bspw. durch Bündelung von wichtigen Titeln mit weniger wichtigen in elektronischen Abonnements oder Zwangsabonnements gedruckter Ausgaben zusätzlich zum Onlinebezug).<sup>15</sup> Bibliotheken setzen dieser geballten Marktmacht Konsortien entgegen, um ihre Interessen besser wahren zu können.<sup>16</sup> Die hohen Kosten, die der öffentlichen Hand durch diese starke Position der Verlage entstehen, sind umso weniger vertretbar, als die Haupttätigkeiten, die zur Veröffentlichung eines Artikels führen – Forschung, Verfassen des Textes

---

<sup>12</sup>Bereits 1993 konstatierten Grötschel u. a. (1993), 2: „Nun ist der Bereich des Publizierens ein zweites Mal von der elektronischen Revolution betroffen. Heute sind die technischen Mittel, zu vervielfältigen und professionell zu veröffentlichen, dramatisch preiswerter geworden. Sie stehen damit zunehmend auch den Erstelern [sic!] der zu publizierenden Information zur Verfügung.“ Vgl. auch Hagenhoff u. a. (2007), 15 u. Oppenheim (2008), 579.

<sup>13</sup>Vgl. exemplarisch Kölbl (2011), 206; Hagenhoff u. a. (2007), 10ff.; Kirchgässner (2003).

<sup>14</sup>White u. Creaser (2007), 17. Weitere Zahlen und Grafiken zu den Preissteigerungen vgl. auch Allianz der deutschen Wissenschaftsorganisationen (2011b), 21f.

<sup>15</sup>Vgl. z. B. Wiesner (2004), 1 u. 7.

<sup>16</sup>Vgl. Kölbl (2011), 206. Vgl. auch exemplarisch Orsdel u. Born (2009) für eine Situationsbeschreibung der Verhandlungen mit Verlagen in den USA 2009.

und Peer-Review-Verfahren –, Professoren und wissenschaftliche Mitarbeiter leisten, die nicht von den Verlagen bezahlt werden, sondern von den Universitäten und Forschungseinrichtungen, den Hauptabnehmern der Zeitschriften.

Die Aufbau eigener, digitaler Publikationsplattformen durch Wissenschaftseinrichtungen, die ohnedies bereits über eine ausgeprägte kommunikationstechnologische Infrastruktur verfügen<sup>17</sup>, scheint daher eine Lösung zu bieten, sich zumindest zum Teil von der Abhängigkeit durch die Verlage zu befreien. Die vom *Open Society Institute* 2001/2002 ins Leben gerufene *Budapest Open Access Initiative* (BOAI) begründet ihren Aufruf zu diesem als *Open Access* benannten Modell zudem mit dem Abbau von Zugangsbeschränkungen zu Wissen sowie einer grundsätzlich größeren Nutzbarkeit von Publikationen für Forschung, Ausbildung und Entwicklung weltweit, die für Arm und Reich über die Wissenschaften hinaus positive Auswirkungen haben wird.<sup>18</sup> Die darauf folgende *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* von 2003 will das Internet nutzen „[...] as a functional instrument for a global scientific knowledge base and human reflection [...]“ und fordert zur Erlangung dieses Ziels: „Content and software tools must be openly accessible and compatible.“ Open Access müsse in diesem Sinne folgende Bedingungen erfüllen:

- The author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly [...], subject to proper attribution of authorship [...].
- A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus published) in at least one online repository [...].<sup>19</sup>

Die Veröffentlichung z. B. eines Zeitschriftenartikels hat aber nicht nur ökonomische, technische oder wissenschaftsfördernde Aspekte – für einen Wissenschaftler kann die Veröffentlichung eines Artikels in einer Verlagszeitschrift mit hohem Impact Factor und hoher Reputation existenziell sein und beispielsweise über seine Berufung als Professor entscheiden.<sup>20</sup> Die traditionelle wissenschaftliche Veröffentlichungspraxis ist somit eng verbunden mit dem internationalen Wissenschaftssystem und nicht leicht zu reformieren.

---

<sup>17</sup>Vgl. Hagenhoff u. a. (2007), 14.

<sup>18</sup>BOAI (2002).

<sup>19</sup>Berlin Declaration on Open Access (2003). Diese Definition findet sich wortgleich im IFLA (2004).

<sup>20</sup>Vgl. z. B. Hagenhoff u. a. (2007), 1 u. Wiesner u. a. (2004), 175.

Neue Open-Access-Zeitschriften müssen ebenso konsistente Peer-Review-Kriterien anwenden, um dem Vergleich mit traditionellen Publikationsorganen standhalten und eine ähnlich hohe Bedeutung aufbauen zu können. Dies ist aber in kurzer Frist nicht zu erreichen. Es gibt daher verschiedene Strategien, die Anzahl von Open-Access-Publikationen zu erhöhen. Die [BOAI](#) definiert zwei Wege, um offenen Zugang zu wissenschaftlicher Literatur zu erreichen:

1. Self-Archiving: First, scholars need the tools and assistance to deposit their refereed journal articles in open electronic archives, a practice commonly called, self-archiving. When these archives conform to standards created by the Open Archives Initiative, then search engines and other tools can treat the separate archives as one. Users then need not know which archives exist or where they are located in order to find and make use of their contents.
2. Open-access Journals: Second, scholars need the means to launch a new generation of journals committed to open access, and to help existing journals that elect to make the transition to open access. Because journal articles should be disseminated as widely as possible, these new journals will no longer invoke copyright to restrict access to and use of the material they publish. [...] Because price is a barrier to access, these new journals will not charge subscription or access fees, and will turn to other methods for covering their expenses. [...] <sup>21</sup>

Im Folgenden sollen diese beiden Strategien, auch der „goldene“ und „grüne“ Weg zu Open Access genannt, näher erläutert werden. Besonderes Augenmerk wird dabei auf die Selbstarchivierung gelegt, da diese im Fokus dieser Arbeit steht.

### **2.1.1 Der goldene Weg des Open Access: Alternative Publikationsformen**

Ab den 90er Jahren des 20. Jahrhunderts wurden an Universitäten und anderen Institutionen die ersten elektronischen Zeitschriften gegründet, die durch ein Begutachtungsverfahren geprüfte Artikel kostenfrei im Internet veröffentlichten. Im Prinzip wird dabei das althergebrachte Modell der Veröffentlichung wissenschaftlicher Zeitschriftenartikel beibehalten – Einreichung, Begutachtung, Verbesserung durch den Autor, Wiedereinreichung, Annahme, Publikation als abgeschlossener statischer Text. Da der Text aber

---

<sup>21</sup>[BOAI \(2002\)](#). Erste Appelle zur Selbstarchivierung stammen allerdings bereits aus der Zeit vor der Einrichtung von Repositorien: Stevan Harnad veröffentlichte im Juli 1994 einen „Subversive proposal“, der Wissenschaftlern empfahl, ihre Veröffentlichungen aus Peer-Review-Zeitschriften auf öffentlichen Servern frei zur Verfügung zu stellen (vgl. [Harnad \(1995\)](#)).



nun nur elektronisch erscheint, ist keine Verlagsinfrastruktur für Druck und Vertrieb mehr notwendig. Die digitale Infrastruktur stellen die der Institution angehörenden Rechenzentren zur Verfügung, die Redaktion wird vollständig von Wissenschaftlern oder auch Bibliothekaren gestellt.

Diese Art der Veröffentlichung nennt man den „goldenen Weg“ des Open Access, da die Publikationen originär frei für jeden zur Verfügung gestellt werden. Dies muss allerdings nicht unbedingt heißen, dass kein Geld für die Publizierung gezahlt wurde: Es gibt auch als „hybrides Open Access“ bezeichnete Geschäftsmodelle kommerzieller Verlage<sup>22</sup>, die die Zahlung einer Gebühr durch die Autoren vorsehen, zu deren Unterstützung mittlerweile ein entsprechender Fonds durch die *Deutsche Forschungsgemeinschaft (DFG)* eingeführt wurde.<sup>23</sup> Auch haben einige Forschungsinstitutionen Verträge mit Verlagen abgeschlossen, die gegen Zahlung einer vereinbarten Summe pauschal eine Open-Access-Option für die dortigen Publikationen ihrer gesamten Forscher vorsehen.<sup>24</sup>

### 2.1.2 Der grüne Weg des Open Access: Selbstarchivierung

Beinhaltet der goldene Weg demnach die Bereitstellung eines genuin frei verfügbaren Artikels, setzt der grüne Weg des Open Access auf die *Selbstarchivierung* einer gleichzeitig in einer üblichen kommerziellen Zeitschrift erscheinenden oder bereits erschienenen Publikation. Selbstarchivierung bedeutet hierbei das Hochladen einer PDF-Version der Publikation beispielsweise auf die eigene Webseite oder – vor allem – auf einen institutionellen oder disziplinären Dokumentenserver.<sup>25</sup> Die Betreiber dieses als Repositorium bezeichneten Servers, meist Universitätsbibliotheken oder die Rechenzentren von Forschungseinrichtungen, sorgen für die Infrastruktur, die zur langfristigen Sichtbarmachung der Dokumente im Internet notwendig ist, und verfolgen meist auch verbindlich definierte Strategien der Langzeitarchivierung. Die Selbstarchivierung mittels eines Repositoriums anstelle einer Webseite hat somit technische Vorteile und bietet eine Perspektive von Nachhaltigkeit.

Disziplinäre Repositorien oder Fachrepositorien sind Dokumentenserver, die Publikationen zu einem bestimmten Fachgebiet hosten. Auch sie werden oft von einer Forschungsinstitution betrieben, sind aber nicht auf deren Mitglieder beschränkt, sondern

---

<sup>22</sup>Vgl. Krönung u. a. (2010), 232. Dort wird auch das System der Nationallizenzen als „hybrides“ Open-Access-Modell bezeichnet. Nationallizenzen sind nach Meinung des Verf. dieser Arbeit aber nur teilweise mit dem Open-Access-Gedanken vereinbar, da sie nicht alle Forderungen der Budapester und Berliner Erklärung erfüllen.

<sup>23</sup>Vgl. Deutsche Forschungsgemeinschaft (2010).

<sup>24</sup>Vgl. SUB Göttingen (2011).

<sup>25</sup>Vgl. Oppenheim (2008), 579.

offen für jeden aus der Fachcommunity. Ein Beispiel hierfür ist das Social Science Open Access Repository ([SSOAR](http://www.ssoar.info/)) für sozialwissenschaftliche Dokumente, das durch *GESIS – Leibniz-Institut für Sozialwissenschaften* betrieben wird.<sup>26</sup>

Institutionelle Repositorien hingegen haben das Ziel, fachübergreifend die Publikationen aller Mitarbeiter der sie betreibenden Institution zu sammeln und – unter Einhaltung urheberrechtlicher Beschränkungen – öffentlich zur Verfügung zu stellen.

Eine bekannte Definition eines [IR](#) stammt von Clifford Lynch:

[...] a university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution. [...] At any given point in time, an institutional repository will be supported by a set of information technologies, but a key part of the services that comprise an institutional repository is the management of technological changes, and the migration of digital content from one set of technologies to the next as part of the organizational commitment to providing repository services. An institutional repository is not simply a fixed set of software and hardware.<sup>27</sup>

[IR](#) haben also nicht ausschließlich die Funktion eines reinen Datenspeichers, sondern sollen aktiv die Verbreitung ihrer Inhalte fördern sowie für persistente Langzeitverfügbarkeit und -archivierung sorgen, je nach den gültigen und sich weiterentwickelnden technischen Standards. Beispiele für [IR](#) sind: Konstanzer Online-Publikations-System ([KOPS](#)) an der Universität Konstanz oder Zurich Open Repository and Archive ([ZORA](#)) an der Universität Zürich.

Das *Directory of Open Access Repositories* (OpenDOAR) verzeichnet 1.654 [IR](#) weltweit, davon 130 in Deutschland, 8 in der Schweiz und 6 in Österreich. Disziplinäre Repositorien gibt es weltweit 226, davon werden 14 in Deutschland, eines in der Schweiz und eines in Österreich betrieben.<sup>28</sup> Diese Arbeit fokussiert sich vorwiegend auf die Rolle und Funktionen von [IR](#), wegen ihrer großen Überzahl gegenüber disziplinären Repositorien. Viele Ergebnisse sind allerdings übertragbar, da die hauptsächlichen Funktionalitäten übereinstimmen und nur die Trägerschaft – und damit die Beziehung zu den Dokumentablieferern – abweicht.

---

<sup>26</sup>Homepage: <<http://www.ssoar.info/>> (Abruf: 10.08.2011).

<sup>27</sup>Lynch (2003), Abschn. 'Defining Institutional Repositories'.

<sup>28</sup><<http://www.opendoar.org/index.html>, Stand: 01.08.2011.

Bei den Dokumenten, die auf Repositorien archiviert werden, kann es sich um verschiedene Arten von Versionen einer Publikation handeln. Besonders in naturwissenschaftlichen Disziplinen hat beispielsweise der Austausch von *Preprints* bereits vor der Veröffentlichung große Tradition, so ist auch *arXiv*, eines der bekanntesten disziplinären Repositorien, ursprünglich als Preprint-Server für die Physik implementiert worden.<sup>29</sup> Besondere Bedeutung erhält heute der grüne Weg aber durch das Hochladen von *Postprints*, wobei es sich um die Version handelt, die – nach dem Peer-Review-Verfahren – vom Verlag zur Publikation angenommen wurde, aber noch nicht dem Textsatz und der Seitengestaltung der veröffentlichten Version entspricht.<sup>30</sup> Ein Postprint wird auch als „accepted version“ bezeichnet, die darauf folgende Verlagsversion im Gegensatz dazu als „published version“.<sup>31</sup>

Eine Postprint-Version ist somit, nach bisher gültigen wissenschaftlichen Standards, nicht zitierfähig – hierfür wird weiterhin die Verlagsversion benötigt. Daher erlauben viele Verlage, auch wenn sie sich über ein Copyright Transfer Agreement (CTA) die Nutzungsrechte an dem Artikel haben übertragen lassen, eine Selbstarchivierung auf einem IR, wobei das Dokument oft erst nach einer Sperrfrist, einem sog. „Embargo“, von zwölf oder 24 Monaten frei im Internet zur Verfügung gestellt werden darf.<sup>32</sup>

In Deutschland unterstützen auch der Wissenschaftsrat und die DFG die Forderung nach einem durch die Verlage eingeräumten Zweitveröffentlichungsrecht, das dem Autor die öffentlich zugängliche Selbstarchivierung unkompliziert erlauben würde.<sup>33</sup> Mittlerweile habe einige Institutionen Richtlinien verabschiedet, die ihre Wissenschaftler zum Hochladen ihrer Publikationen auf IR auffordert.<sup>34</sup> Grundsätzlich kann bisher aber nicht von einem großen Erfolg des grünen Weges gesprochen werden: In einer statistischen Analyse ermitteln Björk u. a. (2010), dass insgesamt 20,4% aller 2008 erschienenen Artikel als Open Access zur Verfügung stehen, summiert aus 8,4% goldener Weg und 11,9% grüner Weg. Aber nur 2,9% aller Publikationen sind auf IR selbstarchiviert worden, die restlichen auf grünem Wege freigegebenen Publikationen befinden sich entweder in

---

<sup>29</sup>Vgl. Kölbl (2011), 184; Hagenhoff u. a. (2007), 29ff. Homepage von *arXiv*: <<http://arxiv.org/>> (Abruf: 10.08.2011).

<sup>30</sup>Zur Definition der Begriffe Pre- und Postprint vgl. auch die Informationen auf den SHERPA/RoMEO-Seiten, einer Datenbank, die Auskunft darüber gibt, welche Zeitschriftenverlage Selbstarchivierung erlauben: SHERPA RoMEO (2006). Vgl. auch Hagenhoff u. a. (2007), 14f.

<sup>31</sup>Vgl. z. B. Springer (2011b).

<sup>32</sup>Vgl. z. B. Springer (2011b).

<sup>33</sup>Vgl. Wissenschaftsrat (2001), 26.

<sup>34</sup>Z. B. die Universität Zürich (vgl. Univ. Zürich (2009)) oder die amerikanischen National Institutes of Health (vgl. US National Institutes of Health (2010)).

Fachrepositorien (5,1%) oder auf persönlichen Webseiten (3,9%).<sup>35</sup> Müller u. a. (2009) konstatieren, dass IR u. a. deswegen bisher keinen großen Erfolg haben, weil sie den Wissenschaftlern keinen Mehrwert bieten und ihre Inhalte nicht in deren üblichen Rechercheumgebungen präsent sind.<sup>36</sup> Salo (2008) stellt grundsätzlich fest, dass die technische Infrastruktur der Repositorien große Vorteile für Langzeitarchivierung und persistente Verfügbarkeit bietet, dieses aber für Wissenschaftler keine relevanten Services sind. Oft wird in der Forschungsliteratur – gestützt durch qualitative Interviews – zudem die Auffassung vertreten, Wissenschaftler hielten Open Access für nicht wichtig, da sie über die Lizenzierungen ihrer Bibliothek Zugriff auf die relevante Literatur haben.<sup>37</sup>

Im Folgenden soll, u. a. auf der Grundlage neuerer statistischer Erhebungen, die Interessenlage der Wissenschaftler gegenüber Open Access untersucht und damit eine Basis für die Erstellung von Anforderungen geschaffen werden, die für einen größeren Erfolg von IR notwendig erachtet werden.

## 2.2 Anforderungen an Institutionelle Repositorien aus Sicht der Wissenschaft

### 2.2.1 Die Einstellung von Wissenschaftlern zu Open Access und Selbstarchivierung

Institutionelle und auch fachliche Repositorien stellen Publikationen von Forschern für Forscher zur Verfügung. Es ist daher selbstverständlich, dass bei ihrer Evaluierung und Weiterentwicklung die Anforderungen von Wissenschaftlern im Mittelpunkt stehen müssen, nicht die Vorstellungen der sie betreibenden universitären Bibliotheken oder anderer Forschungsinstitutionen. 2005 formulierte ein Positionspapier der DFG: Wissenschaftler erwarten die „[...] Erschließung und Verzeichnung der digitalen Publikationen nach internationalen Standards, ihre Präsentation in fachspezifischer Umgebung und ihre Einbindung in fachspezifische Informationssysteme“ sowie „[...] klar strukturierte und logisch zu bedienende Nachweis- und Suchsysteme, welche sie in wenigen Schritten direkt zu dem

---

<sup>35</sup>Die entsprechenden Tabellen aus Björk u. a. (2010) sind einzeln erhältlich unter <<http://dx.doi.org/10.1371/journal.pone.0011273.t003>> und <<http://dx.doi.org/10.1371/journal.pone.0011273.g003>> (Abruf: 10.08.2011). Der Analyse liegt ein Sample von 1.837 Zeitschriftenartikeln zugrunde, die 2008 erschienen sind. Da die Auswertung im Herbst 2009 erfolgte, sind für den grünen Weg somit nur Artikel berücksichtigt, deren Selbstarchivierung keinem oder einem höchstens einjährigem Embargo unterlag, die Anzahl müsste bei Berücksichtigung der einem zweijährigen Embargo unterliegenden Artikel daher höher sein.

<sup>36</sup>Müller u. a. (2009), Abschn. 2.

<sup>37</sup>Z. B. Davis u. Connolly (2007), Abschn. ‘Access a non-issue for faculty’.

gewünschten Dokument führen“.<sup>38</sup> Repositorien können nur durch die Bereitstellung von innovativen und effizienten Services und durch das Eingehen auf die Anforderungen der Wissenschaftsprozesse Unterstützer finden. Es muss daher eine genaue Analyse auf der Grundlage von unter Wissenschaftlern erhobenen empirischen Daten stattfinden, unter Einbeziehung der aktuellen Veränderungen der Wissenschaftskommunikation.

Zwei Projekte untersuchen mit Hilfe von Umfragen die Einstellungen von Wissenschaftlern zu Open Access, um hieraus Erkenntnisse darüber zu gewinnen, wie der goldene und grüne Weg in den Fachcommunities mit größerem Erfolg bekannt gemacht werden kann.

- Das Projekt „Study of Open Access Publishing“ (SOAP) veranstaltete 2010 eine umfangreiche Umfrage zu Publikationserfahrungen in Open-Access-Zeitschriften und erhob dabei auch Daten zu allgemeinen Ansichten und Einstellungen der Wissenschaftler gegenüber der frei zugänglichen Verfügbarkeit von Forschungsliteratur im Internet.<sup>39</sup>
- „Publishing and the Ecology of European Research“ (PEER) analysiert, ebenso wie SOAP unter Beteiligung des Verlagswesens, den Einfluss des grünen Weges auf das europäische System der Wissenschaften (Sichtbarkeit von Autoren, Zugangsmöglichkeiten zu Publikationen, Auswirkungen auf wissenschaftliche Verlagszeitschriften). In diesem Rahmen wurde auch eine Umfrage unter Wissenschaftlern zu ihrem Publikations- und Rechercheverhalten sowie zum Umgang mit Open-Access-Repositorien durchgeführt.<sup>40</sup>

Im Folgenden schließen sich die Zusammenfassung einiger Ergebnisse und Schlussfolgerungen an, die daraus für den Betrieb von IR gezogen werden können.

### Allgemeine Einstellung zu Open Access

- 89% aller teilnehmenden Wissenschaftler glauben, dass die Veröffentlichung von Artikeln in Open-Access-Zeitschriften von Nutzen für das eigene Fachgebiet ist.

---

<sup>38</sup>Deutsche Forschungsgemeinschaft (2005), 11.

<sup>39</sup>Insgesamt wurden die Antworten von 38.538 Wissenschaftlern aus der ganzen Welt ausgewertet, die in den fünf der Umfrage vorausgehenden Jahren mindestens einmal in einem Peer-Review-Journal veröffentlicht haben (vgl. Dallmeier-Tiessen u. a. (2011a), 3). Homepage des SOAP-Projekts: <<http://project-soap.eu/>> (Abruf: 10.08.2011).

<sup>40</sup>Einladungen zur Umfrage wurden anhand von Autorenlisten der am Projekt beteiligten Verlage versandt; es wurden dann die Antworten von 3.139 Forschern aus dem Gebiet der EU ausgewertet (vgl. Fry u. a. (2009), 19). Homepage des PEER-Projekts: <<http://www.peerproject.eu/home/>> (Abruf: 10.08.2011).

Als Gründe werden u. a. genannt: höherer Nutzen für die Wissenschaft (36%), Forschung ist öffentliches Eigentum (18%), höherer Nutzen für einen selbst (10%).<sup>41</sup>

- Die Varianz der Antworten ist unter den nach offenem bis beschränktem Zugang zu Peer-Review-Zeitschriften eingeteilten Gruppen nur leicht abweichend. Auch Wissenschaftler von Institutionen, die über großzügigen Zugang zu lizenzierten elektronischen Ressourcen verfügen, sehen Open Access also positiv.<sup>42</sup>
  - 66,4% verbinden mit dem Begriff Open Access „Online/electronic format“, 78,7% „free access“ und 66,3% „fulltext available“ – aber nur 6% „self archiving“.<sup>43</sup>
- ⇒ Wissenschaftler sehen Open Access grundsätzlich positiv, unabhängig davon, ob sie als Rezipienten davon Gebrauch machen. Open Access bedeutet für sie grundsätzlich „freier elektronischer Zugang zu Informationen und Publikationen“, einen Unterschied zwischen goldenem und grünem Weg machen sie dabei nicht.

### Einstellungen zu Peer-Review-Zeitschriften

- Für wichtige Faktoren, die für die Veröffentlichung eines Artikels in einem Peer-Review-Journal sprechen, halten 85,2% „Widespread availability of my work“, 80,5% „Esteem/reputation“, 76,1% „Dissemination to target audience“, 69,8% „Peer review (getting my work evaluated)“.<sup>44</sup>
- 95,9% bezeichnen „Peer review journals“ als wichtige Literaturquelle für ihre eigenen Forschungen. Auf den nächsten Plätzen folgen „Book chapter“ (55,6%), „Monographs“ und „Conference proceedings“ (jeweils 47,4%).<sup>45</sup> 58,5% lesen mehr als 50 Artikel aus Peer-Review-Zeitschriften pro Jahr.<sup>46</sup>
- 34,4% der Befragten haben es schon häufiger, immerhin noch 42,8% manchmal erlebt, dass sie keinen direkten Zugriff auf einen relevanten Artikel einer Peer-Review-Zeitschrift hatten<sup>47</sup>, 50,2% würden in diesem Fall eine Open-Access-Quelle

---

<sup>41</sup>Dallmeier-Tiessen u. a. (2011b), Folien 9 u. 16. Vgl. auch Dallmeier-Tiessen u. a. (2011a), 4f.

<sup>42</sup>Dallmeier-Tiessen u. a. (2011b), Folie 13.

<sup>43</sup>Fry u. a. (2009), 82.

<sup>44</sup>Fry u. a. (2009), 85. (Zahlen der Angaben „very important“ und „important“ wurden summiert.)

<sup>45</sup>Fry u. a. (2009), 89. (Zahlen der Angaben „very important“ und „important“ wurden summiert.)

<sup>46</sup>Fry u. a. (2009), 89.

<sup>47</sup>Fry u. a. (2009). Diese Zahlen stehen im Gegensatz zu den Erkenntnissen von Davis u. Connolly (2007), Abschn. ‘Access a non-issue for faculty’, die aussagen, dass Wissenschaftler den Eindruck haben, auf alle relevante Literatur zugreifen zu können. U. U. liegt diese Diskrepanz darin begründet, dass PEER europaweit Wissenschaftler aus verschiedenen Institutionen befragte, während bei Davis u. Connolly (2007) nur eine kleine homogene Gruppe aus einer Institution mit identischen (und anscheinend guten) Zugangsbedingungen zugrunde liegt.

suchen, um die Publikation zu erreichen, weitere zweistellige Prozent würden einen Kollegen befragen (22,9%) oder die Bibliothek bemühen (19,5%) – aber nur 3,7% den Artikel kostenpflichtig erwerben<sup>48</sup>.

- Neben Befürchtungen, das Copyright (47,6%) oder vom Verlag gesetzte Embargo-Beschränkungen (40,6%) zu verletzen, ist das wichtigste Gegenargument gegen die Bereitstellung einer Publikation in einem IR: „Reluctant to put my research publications in a repository where other materials have not been peer-reviewed“ (48,2%).<sup>49</sup>
- ⇒ Peer-Review-Zeitschriften stellen mit großem Abstand die wichtigste Informations- und Literaturquelle für Forscher dar und gelten als effizientestes Verbreitungsmedium der eigenen Forschungsergebnisse, denen sie zusätzlich Seriosität und Reputation verleihen. In Repositorien sollten Peer-Review-Zeitschriften deutlich als Quelle eines hochgeladenen Artikels erkennbar und ein Link zur Homepage der Zeitschrift und zur Verlagsversion vorhanden sein. Je mehr Dokumente als begutachtet erkennbar sind, desto höher steigt die Reputation des Repositoriums und seine Attraktivität für Wissenschaftler.<sup>50</sup> Kostenpflichtige Angebote von Verlagen über die Lizenzierungen der eigenen Institutionen hinaus werden kaum angenommen, d. h. dass Open Access grundsätzlich die Wahrscheinlichkeit der Rezeption eines Artikels steigern sollte.

## Erfahrung mit Repositorien

- 23,6% der Befragten haben bereits einmal ein Dokument in ein IR hochgeladen, 24,4% in ein Fachrepositorium. 45% taten dies zumindest einmal aus eigenem Antrieb, also ohne ausdrückliches Mandat ihrer Institution.<sup>51</sup>
- Der freie Zugang für alle (85,6%) bzw. für Wissenschaftler mit beschränktem Zugang zu Peer-Review-Zeitschriften (81,5%) ist mit das wichtigste Motiv, eigene Publikationen in Repositorien hochzuladen. Weitere Gründe sind: „Widespread availability of my research“ (88,6%), „Speed of dissemination“ (69,7%), „Possibility of increased citations to the output“ (68,4%).<sup>52</sup>

---

<sup>48</sup>Fry u. a. (2009), 90f.

<sup>49</sup>Fry u. a. (2009), 88.

<sup>50</sup>Vgl. hierzu auch Davis u. Connolly (2007), Abschn. ‘Reasons for not using a digital repository’.

<sup>51</sup>Fry u. a. (2009), 86f.

<sup>52</sup>Fry u. a. (2009), 87. (Zahlen der Angaben „very important“ und „important“ wurden summiert.)



- 46,1% der Befragten bevorzugen ein Fachrepositorium zum Hochladen eigener Publikationen, 22,1% ein IR und 8,3% ihre eigene Webseite.<sup>53</sup>

⇒ Die Verbreitung, Rezeption und in Folge Zitation der eigenen Forschungsergebnisse stellen die wichtigsten Motive auch für das Hochladen von Dokumenten in IR dar. Fachliche Repositorien werden allerdings deutlich bevorzugt, die Nähe zur eigenen Disziplin ist wichtiger als zur Institution.<sup>54</sup> Publikationen aus IR müssen daher auch an den Orten auffindbar sein, an denen fachlich gesucht wird (in Fachportalen, -bibliographien und -repositorien<sup>55</sup>), aber auch dort, wo grundsätzlich Internetrecherchen starten: über Suchmaschinen. Wichtig hierzu: eine eindeutige, übergeordnete Identifizierung der Autoren, des Fachgebiets, der Peer-Review-Zeitschrift, der Affiliation/Institution. Repositorien dürfen keine „Inseln“<sup>56</sup>, sondern müssen vollständig in die üblichen Recherchestrategien integrierbar sein. Objekte in IR sollten sachlich daher in einem Maße erschlossen sein, das eine Präsentation und ein Auffinden auch in fachlichen Zusammenhängen erlaubt. Eine Integration der Metadaten in die OPACs der Bibliotheken (und damit oft in regionale und überregionale Verbundkataloge) wird von beinahe der Hälfte der IR geleistet<sup>57</sup>, hat allerdings mehr Relevanz für die lokalen Bibliotheken und Forschungsinstitutionen als für die Wissenschaftler.<sup>58</sup>

## Suchgewohnheiten

- Um relevante Literatur zu finden, nutzen normalerweise 74,7% der Befragten „Bibliographies and citation lists“, 53% „Bibliographic databases“ und 52% „subject-based portals or repositories“. *Google* und andere Suchmaschinen werden in die-

---

<sup>53</sup>Fry u. a. (2009), 88.

<sup>54</sup>Vgl. auch Foster u. Gibbons (2005), Abschn. ‘What Faculty Members Want’. Die Autoren stellten in Interviews zudem fest, dass der Begriff „institutionell“ Wissenschaftlern den Eindruck vermittelt, als werde ihre individuelle Forschungsleistung gegenüber der allgemeinen der Institution niedriger bewertet.

<sup>55</sup>Mit der Einbindung der in IR vorhandenen Publikationen in disziplinäre Repositorien und Fachbibliographien befasst sich das DFG-geförderte Projekt „Open-Access-Fachrepositorien“ (OAFR) an der UB Konstanz, Homepage: <<http://www.ub.uni-konstanz.de/bibliothek/projekte/open-access-fachrepositorien.html>> (Abruf: 10.08.2011). Enge Kooperationen zwischen IR und Fachrepositorien fordert auch Salo (2008), 120.

<sup>56</sup>McKay (2007), Abschn. ‘Information seekers’.

<sup>57</sup>Laut einer im DRIVER-Projekt durchgeführten Umfrage bringen 48% der IR ihre Metadaten in die zugehörigen Bibliothekskataloge (vgl. Graaf (2009), 86).

<sup>58</sup>Laut Fry u. a. (2009), 90, nutzen 65,1% den OPAC kaum oder nie, um für sie relevante Literatur zu finden. Institutionelle Nachweise sind anscheinend nicht entscheidend. Ob die Suchmöglichkeiten der IR selbst zur Suche genutzt werden, wurde nicht gefragt. S. a. nächsten Punkt „Suchgewohnheiten“)



sem Zusammenhang von 33,4% genutzt, *Google Scholar* von 25,8% – OAIster, eine Meta-Suchmaschine für Repositorien, nur von 0,3%.<sup>59</sup>

⇒ Auch hier wird deutlich: Fachliche Zusammenhänge sind wichtige Kriterien bei der Literaturrecherche und damit für die Sichtbarkeit einer Publikation. IR werden unmittelbar nicht als Quelle von Literatur gesehen. Grundsätzlich ermöglicht eine genauere und standardisiertere Erschließung auch bessere Suchergebnisse in allgemeinen Suchmaschinen, dies gilt zukünftig in noch höherem Maße, wenn Semantic-Web-Technologien weiter verbreitet sind.

### 2.2.2 Individuelle Mehrwertdienste für Wissenschaftler

Eine explizite repräsentative Umfrage unter Wissenschaftlern über Dienstleistungen, deren Bereitstellung sie durch IR erwarten bzw. begrüßen würden, wurde bisher nicht durchgeführt. Entsprechende Vorschläge können sich allerdings auf Schlussfolgerungen aus allgemeineren Datensätzen wie den oben zitierten und auf qualitative Interviews mit kleineren Forschergruppen stützen.

**Literaturlisten, Wissenschaftlerportfolios** Wissenschaftler pflegen heute eigene Webseiten mit teilweise umfangreichen Publikationslisten, da diese ihren wissenschaftlichen Erfolg belegen. Eine automatische Generierung solcher Listen aus Repositorien inklusive der Links zu Volltexten würde den Aufbau und die Pflege solcher Listen stark vereinfachen. Sie könnten somit Teil eines öffentlich zugänglichen Wissenschaftler-Portfolios sein, das auch Lebenslauf, Forschungsinteressen, Kooperationen etc. enthalten könnte.<sup>60</sup> Hierzu muss es allerdings möglich sein, Publikationen verschiedener IR mit einzubeziehen, da Wissenschaftler im Laufe ihrer Karriere die Institution wechseln können.

**Projektpublikationen** Repositorien können auch die Publikationsplattform für Reports, Deliverables und Abschlussberichte eines Forschungsprojekts sein. Somit wäre eine konsistente Verfügbarkeit und zudem Langzeitarchivierung gewährleistet. Projekte müssen zu diesem Zweck in den Metadaten eindeutig identifizierbar sein.

---

<sup>59</sup>Fry u. a. (2009), 90. Interessant wäre es, diese Frage mit dem Alter der Befragten zu korrelieren: Die Vermutung liegt nahe, dass die älteren Wissenschaftler einen größeren Hang zu traditionellen fachlichen bibliographischen Datenbanken haben, während bei den jüngeren vermutlich ein allgemeiner Trend zu allgemeinen und wissenschaftlichen Suchmaschinen wie *Google* und *Google Scholar* auszumachen wäre (vgl. Deutsche Forschungsgemeinschaft (2005), 5). Deren Bedeutung würde zukünftig dann noch weiter steigen.

<sup>60</sup>Vgl. Foster u. Gibbons (2005), Abschn. 'Enhancing the IR to Meet the Needs of Faculty Users'; Bruch u. Lengenfelder (2009), 5; Fry u. a. (2009), 59.

**Alerting-Dienste, RSS-Feeds** Traditionelle Informationsmethoden über neuerschienene Literatur – der Blick in Zeitschrifteninhaltsverzeichnisse oder Bibliographien – werden immer mehr abgelöst durch individuellere Vorgehensweisen: Recherchen nach relevanten Sachbegriffen in umfangreichen bibliographischen Datenbanken wie *PubMed* oder *Web of Science* sowie regelmäßiges Überprüfen der Homepages bekannter/wichtiger Forscherkollegen („key people“) oder Arbeitsgruppen.<sup>61</sup> Repositorien, die darauf aufsetzenden aggregierenden Anwendungen die Generierung entsprechender Alerting-Dienste oder RSS-Feeds ermöglichen, bieten daher einen Mehrwert, der umso höher ist, als dass Literaturnachweise zwar auch in anderen, kommerziellen Angeboten zu finden sind, im Gegensatz dazu hier aber auch ein Link zum für jeden frei zugänglichen Volltext geboten würde.<sup>62</sup>

**Nutzungsstatistiken, Zitationshäufigkeit** Repositorien können ihren Submittern Statistiken zur Verfügung stellen, wie oft ein Artikel heruntergeladen oder wie oft die Metadaten aufgerufen wurden. Die Nutzungshäufigkeit eines einzelnen Artikels ist zusätzlich zum Impact Factor, der sich auf die ganze Zeitschrift bezieht<sup>63</sup>, als Maßstab für seine Relevanz verwendbar.<sup>64</sup> Zudem können Publikationen aus verschiedenen [IR](#) daraufhin analysiert werden, wie oft andere Publikationen sie zitieren.<sup>65</sup>

**Unterstützung des Publikationsprozesses und des Peer Reviews** [IR](#) können auch eine Rolle spielen im Entstehungs- und Veröffentlichungsprozess eines Artikels. Bei der Kooperation mehrerer Autoren kann durch eine Versionierung die gemeinsame Bearbeitung eines Dokuments den bisherigen Austausch per Email ablösen. Autoren können bereits vor Einreichen des Artikels bei einer Zeitschrift mit dem [IR](#) als technischer Plattform Kollegen und Spezialisten aus der Fachcommunity einladen, Annotationen hinzuzufügen oder bestimmte noch unklare Fragestellungen gemeinsam zu klären. Auch nach der Veröffentlichung könnten Annotationstools bei Postprints noch eine fachliche Diskussion in Gang setzen. Für den offiziellen Reviewprozess, der oft mehrere neue Versionen eines Artikels zur Folge hat, und für

---

<sup>61</sup>Fry u. a. (2009), 62.

<sup>62</sup>Fry u. a. (2009), 59; Boosen (2008), 59.

<sup>63</sup>Vgl. Boosen (2008), 58

<sup>64</sup>Das von der DFG geförderte Projekt „Open-Access-Statistik“ (2008 – 2013) erarbeitet Standards und die technischen Grundlagen sowie Services für die Erhebung von Zugriffsstatistiken, Homepage: <http://www.dini.de/projekte/oa-statistik/>.

<sup>65</sup>Dies ist Thema eines weiteren von der DFG geförderten Projektes, „Distributed Open-Access Reference Citation Services (DOARC)“, Homepage: <http://www.isn-oldenburg.de/projects/doarc2/> (Abruf: 10.08.2011).

das endgültige Einreichverfahren könnten Autoren, Verlage und die als Gutachter bestimmten Personen das IR ebenfalls nutzen, unter Vereinbarung gemeinsamer Standards (Dateiformate, Authentifizierung, Schnittstellen für APIs etc.).<sup>66</sup>

### 2.2.3 Kollaborative Mehrwertdienste im Rahmen neuer Formen der Wissenschaftskommunikation

IR können aber nicht nur individuelle Dienstleistungen für Wissenschaftler erbringen, sie haben auch das Potenzial, eine wichtige Rolle bei allgemeinen, übergeordneten Entwicklungen im Wissenschaftsprozess einzunehmen.

Die Formen der wissenschaftlichen Kommunikation und der Wissensgenerierung haben sich in den letzten Jahrzehnten durch die Möglichkeiten digitaler Medien und Verbreitungsformen verändert: Der klassische, lineare Prozess umfasst das Forschen einer individuellen Person oder einer lokal situierten Arbeitsgruppe, u. U. die Präsentation und Diskussion erster Resultate auf Konferenzen und schließlich eine Veröffentlichung in Form meist eines Zeitschriftenartikels, der dann die Grundlage für Verbreitung und Rezeption der Ergebnisse ist. Moderne Technologien ermöglichen nun neue Formen der Rezeption, Kooperation und Kollaboration: Bereits Preprints können von Fachkollegen mit wenigen Klicks zur Kenntnis genommen werden, Forschungsrohdaten parallel zur Publikation digital zur Verfügung gestellt werden, global agierende Arbeitsgruppen gemeinsam und online Dokumente verfassen.

Eng miteinander verwandte Schlagwörter dieser neuen Formen des Wissenschaftsprozesses sind *eScience*, *cyberinfrastructure*, *collaboration environment*, das Kunstwort *collaboratory*<sup>67</sup> sowie *virtuelle Forschungsumgebungen* und *Forschungsinfrastrukturen*. Sie umfassen

- das integrierte Arbeiten mit den verschiedensten Arten von Ressourcen – zum Teil umfangreiche Primärdaten wie Videos, Audiofiles, Mitschriften von Interviews, statistische Daten als auch Sekundärdaten wie Publikationen –, die durch entsprechende Hard- und Software leicht und unkompliziert global zur Verfügung gestellt werden (*Rechennetze*, *grid computing*)
- die hierdurch mögliche Zusammenarbeit von geographisch verstreuten Wissenschaftlern verschiedener Institutionen (*networked science*, im Gegensatz zu *team*)

---

<sup>66</sup>Foster u. Gibbons (2005), Abschn. ‘What Faculty Members Want’; Salo (2008), 103; Boosen (2008), 57f.

<sup>67</sup>Vgl. Vaart (2010).

*science*, die lokal an einem Forschungszentrum angesiedelte Arbeitsgruppen von Wissenschaftlern unterschiedlicher Herkunft vorsieht)

- die Bereitstellung von Tools, die semantische Verknüpfungen zwischen den verschiedenen Daten zur Verfügung stellen und vernetzte Kollaborationen sowie soziale Interaktionen verteilt arbeitender Wissenschaftler ermöglichen (Wissensvernetzung, *knowledge management* bzw. *networking*).<sup>68</sup>

Die Schwerpunktinitiative „Digitale Information“ der *Allianz der deutschen Wissenschaftsorganisationen*, die eine Verbesserung der Informationsinfrastrukturen, leichteren Zugang zu digital vorliegenden Forschungspublikationen und deren Integration in die tägliche Arbeitsumgebung von Wissenschaftlern erreichen will<sup>69</sup>, definiert eine virtuelle Arbeitsumgebung folgendermaßen:

Eine virtuelle Forschungsumgebung ist eine Arbeitsplattform, die eine kooperative Forschungstätigkeit durch mehrere Wissenschaftler an unterschiedlichen Orten zu gleicher Zeit ohne Einschränkungen ermöglicht. Inhaltlich unterstützt sie potentiell den gesamten Forschungsprozess – von der Erhebung, der Diskussion und weiteren Bearbeitung der Daten bis zur Publikation der Ergebnisse – während sie technologisch vor allem auf Softwarediensten und Kommunikations-netzwerken [sic!] basiert. Virtuelle Forschungsumgebungen sind wesentliche Komponenten moderner Forschungsinfrastrukturen.<sup>70</sup>

IR spielen in diesen Szenarien eine wichtige Rolle, da sie den freien und globalen Zugang zu digitalen Publikationen zur Verfügung stellen sollen. Als dedizierte Dokumentenserver sind sie zudem prädestiniert für zusätzliche Funktionalitäten wie das Management verschiedener Versionen oder den authentifizierten schreibenden Dokumentzugriff kollaborierender Forscher. Zusätzlich zu diesem erweiterten Funktionsumfang ist dabei die Vernetzung der IR entscheidend, die sie aus dem begrenzten institutionellen Charakter heraushebt und damit zum Teil einer virtuellen Forschungsumgebung macht.<sup>71</sup> Mit dem Ziel des Aufbaus moderner Forschungsinfrastrukturen unterstützt daher die Aktionslinie „Repositorien – Ausbau und Entwicklung vernetzter Informationsdienstleistungen

---

<sup>68</sup>Vgl. Neuroth (2009), Folien 6–11 u. 19–20; Oßwald (2008), 516–519; Weller u. a. (2007), 33f.

<sup>69</sup>. Allianz der deutschen Wissenschaftsorganisationen (2011c). Eine entsprechende Initiative existiert ebenfalls auf EU-Ebene: Das European Strategy Forum on Research Infrastructures (ESFRI), dessen Delegierte von den Wissenschaftsministern der Mitgliedsländer ernannt werden, will durch die Konzeption von Roadmaps und die Unterstützung einschlägiger Projekte Forschungsinfrastrukturen stärken, um europäische Forschung international wettbewerbsfähig zu erhalten (vgl. European Commission (2011)).

<sup>70</sup>Allianz der deutschen Wissenschaftsorganisationen (2011a).

<sup>71</sup>Vgl. Reimer (2011), Folie 18: „Repository + advanced search + annotation + collaborative elements = VRE“.

für die Wissenschaft“ des DFG-Förderprogramms „Elektronische Publikationen“<sup>72</sup> u. a. folgende Themenbereiche:

**Interoperabilität, übergreifende Dienstleistungen** Metadaten und Objekte müssen nach Standards zur Verfügung gestellt werden, die auch anderen Diensten erlauben, auf sie zuzugreifen. Gefördert wird in diesem Bereich beispielsweise das Projekt „Persönliche Publikationslisten als hochschulweiter Dienst (PubLister)“ an der UB Bielefeld.<sup>73</sup>

**Automatische Erschließung** Es sollen Tools entwickelt werden, die die in Repositorien vorhandenen Volltexte übergreifend automatisiert nach Kriterien der inhaltlichen Erschließung und Volltextindizierung zur Verfügung stellen. Ein Projekt in diesem Zusammenhang ist „Automatische Anreicherung von OAI-Metadaten mit Hilfe computerlinguistischer Verfahren und Entwicklung von Services für die inhaltsorientierte Vernetzung von Repositorien“ an den Universitäten Bielefeld und Leipzig.<sup>74</sup>

**Kooperation** Institutionelle und fachliche Repositorien sollen eng zusammenarbeiten, auch im internationalen Kontext. Dies bezieht sich auf den Austausch und die Aggregation sowohl von Metadaten als auch statistischen Daten zu Nutzung und Zitationen. Hierzu wird beispielsweise das Projekt „Aufbau eines Netzwerkes zertifizierter Open Access Repositories“ an den Universitäten Osnabrück, Göttingen und Berlin (HU) gefördert.<sup>75</sup>

Auch für die Form der Veröffentlichung der Forschungsergebnisse stehen mittlerweile neue technische Möglichkeiten zur Verfügung: Eine moderne „enhanced publication“ erlaubt es, unmittelbar auf die verschiedenen Informationsobjekte wie statistische Datensätze oder Interviews, auf denen die Publikation basiert, zuzugreifen und damit den Forschungsprozess selbst nachzuvollziehen.<sup>76</sup> Publikationen müssen auch nicht mehr statisch

---

<sup>72</sup>Deutsche Forschungsgemeinschaft (2009).

<sup>73</sup>Homepage: <<http://www.ub.uni-bielefeld.de/biblio/projects/publist.htm>>; s. a. <<http://gepris.dfg.de/gepris/OCTOPUS/?module=gepris&task=showDetail&context=projekt&id=95126119>> (Abruf: 10.08.2011).

<sup>74</sup>Homepage: <[http://www.ub.uni-bielefeld.de/biblio/projects/oai\\_projekt.htm](http://www.ub.uni-bielefeld.de/biblio/projects/oai_projekt.htm)>; s. a. <<http://gepris.dfg.de/gepris/OCTOPUS/?module=gepris&task=showDetail&context=projekt&id=94533356>> (Abruf: 10.08.2011).

<sup>75</sup>Homepage: <<http://www.dini.de/projekte/oa-netzwerk/>>, s. a. <<http://gepris.dfg.de/gepris/OCTOPUS/?module=gepris&task=showDetail&context=projekt&id=41222188>> (Abruf: 10.08.2011). Das Projekt umfasst die Unterprojekte „Distributed Open Access Reference Citation Services“ (DOARC) (s. Fußnote 65 und „Open-Access-Statistik“ (OA-Statistik) (s. a. Fußnote 64).

<sup>76</sup>Vgl. Hogenaar (2009).

bleiben, sondern vermögen sich, bei Einbindung moderner Infrastrukturen in Repositorien, durch fortgesetzte Diskussionen oder Annotationen von Fachkollegen transparent und nachvollziehbar weiterzuentwickeln. Für den Austausch kann in diesem Fall das Austauschprotokoll *Open Archives Initiative Object Reuse and Exchange* (OAI-ORE) genutzt werden, das die Beschreibung von Verknüpfungen zwischen zusammengehörenden einzelnen Objekten einer Publikation transportieren kann.<sup>77</sup>

#### 2.2.4 Die „Thesen zur Zukunft von Open-Access-Repositorien“

2002 gründeten verschiedene, wissenschaftliche Bibliotheken und Rechenzentren vertretende Organisationen die Deutsche Initiative für Netzwerkinformation (DINI), die zum Ziel hat, moderne Kommunikations- und Informationstechnologien an Hochschulen durch die Entwicklung von Standards und Best Practices sowie die Verbesserung des Erfahrungsaustausches zu fördern und dabei ebenfalls eine gemeinsame Strategie für Repositorien in Bezug auf Funktionalitäten, Schnittstellen, Metadatenstandards und Usability entwickelt.<sup>78</sup> Bei Erfüllung eines regelmäßig weiterentwickelten Anforderungskatalogs erhalten Repositorien zudem ein sogenanntes *DINI-Zertifikat* (näheres hierzu s. ab Kap. 3.2). Im November/Dezember 2010 diskutierte ein gemeinsam von der DINI und der Helmholtz-Gesellschaft veranstalteter Workshop „Repositorien – Praxis und Vision“ sechs Thesen zur Zukunft von Open-Access-Repositorien. Diese fassen komprimiert die Ansprüche zusammen, denen Repositorien sich stellen müssen, wenn sie zukünftig einen relevanten Platz in der globalen und sich – wie oben dargestellt – einschneidend verändernden Forschungslandschaft einnehmen wollen:

1. **Repositorien dienen als digitale Schaufenster der Wissenschaft.** Open Access macht Publikationen sichtbarer und erhöht ihre Rezeption.
2. **Nur disziplinäre Angebote werden als Rechercheinstrumente in der Wissenschaft angenommen.** Ein sachlicher Zugang zu den in Repositorien vorhandenen Dokumenten ist unabdingbar.
3. **Repositorien sind nur als aktive und integrative Dienste für Forschung und Lehre attraktiv.** Repositorien müssen Schnittstellen bieten, um ihre Inhalte für andere Services zur Verfügung stellen zu können.

---

<sup>77</sup>Zu OAI-ORE s. beispielsweise Foulonneau u. André (2008), 37ff. Die Spezifikation ist unter <<http://www.openarchives.org/ore/>> erhältlich (Abruf: 10.08.2011).

<sup>78</sup>Vgl. DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 5; DINI (2011).

4. **Repositorien werden als vertrauenswürdige Plattform für das digitale Publizieren wahrgenommen.** Repositorien müssen Standards bei Beschreibung, Langzeitarchivierung und Persistenz der Erreichbarkeit ihrer Objekte einhalten.
5. **Repositorien dienen als Knotenpunkte für eine übergreifende Vernetzung.** Repositorien müssen nicht nur für eine syntaktisch korrekte Bereitstellung über standardisierte Schnittstellen sorgen, sondern auch für semantische Technologien im Rahmen von Linked Open Data bereit sein.
6. **Repositorien dienen als Rückgrat des Forschungsprozesses.** Um eine starke Position im Forschungsprozess einnehmen zu können, müssen sich Repositorien auch als Teil einer virtuellen Forschungsumgebung sehen und neben Publikationen beispielsweise auch Forschungsrohdaten bereitstellen können.<sup>79</sup>

Nicht nur die in den Repositorien vorhandenen Dokumente, auch die sie beschreibenden Metadaten bekommen im Rahmen der Erfüllung dieser Ansprüche im Allgemeinen und der Bereitstellung neuer sowohl individueller wie auch kollaborativer Services im Besonderen eine wichtige Funktion, da sie die notwendige Vernetzung erst möglich machen. Metadaten müssen hierzu anspruchsvolle Anforderungen erfüllen, die über diejenigen eines konventionellen Bibliothekskatalogs hinausgehen: Sie müssen maschinell und von den verschiedensten Tools interpretiert werden können, sie müssen die unterschiedlichsten Arten von Objekten beschreiben können, und sie müssen diese Ansprüche sowohl für umfangreiche aggregierte Sammlungen als auch für den Content nur eines Repositoriums erfüllen.<sup>80</sup> Dies bedeutet, dass die Metadatenelemente der Publikationen nicht nur in ihrer ursprünglichen Umgebung, sondern auch außerhalb eine Bedeutung haben müssen. Bibliotheken, die häufig als Betreiber der Repositorien agieren, haben lange und umfangreiche Erfahrungen in der Erstellung von Metadaten (formale und inhaltliche Erschließung), ihren Strukturen (Verknüpfungen zu Normdaten, Persistenz von Identifiern) und ihrer Präsentation (Recherchetools, Portalsoftware, Fachgateways) und sollten diese Erfahrungen in die oft von ihnen betriebenen Repositorien mit einbringen, um die Entstehung moderner Wissenschaftsumgebungen durch ihre Mittel zu unterstützen.<sup>81</sup>

In den folgenden Kapiteln soll daher die Situation der Metadatenerschließung in [IR](#) analysiert und ein entsprechendes Anforderungsprofil entworfen werden, das die ermittelten Bedürfnisse der modernen Wissenschaftskommunikation berücksichtigt.

---

<sup>79</sup>[Kindling u. a. \(2010\)](#).

<sup>80</sup>[Gold \(2007\)](#), Abschn. 1.4.

<sup>81</sup>Vgl. [Oßwald \(2008\)](#), 519.



## 3 Die Erschließung selbstarchivierter Dokumente in Institutionellen Repositorien

### 3.1 Bedeutung von Metadaten für Elektronische Dokumente

Anders als in konventionellen Bibliothekskatalogen beschreiben Metadaten in [IR](#) beinahe<sup>82</sup> ausschließlich digital vorliegende Dokumente, auf die global oder bei urheberrechtlichen Einschränkungen zumindest institutionsweit zugegriffen werden kann. Die Dokumente – inklusive Titelblatt, eigentlichem Text sowie Referenzen – sind integraler Bestandteil des World Wide Web und können von Suchmaschinen wie *Google* oder *Yahoo* nach den selben Kriterien indexiert und in Suchergebnislisten präsentiert werden wie andere elektronische Ressourcen. Auch ohne eigens erfasste Metadaten sind Dokumente im Internet prinzipiell auffindbar. Der große Vorteil, den Suchmaschinen für Wissenschaftler gegenüber anderen bibliographischen und bibliothekarischen Nachweissystemen haben, ist zudem ihr universeller Charakter, die Indexierung auch unselbständiger Literatur, und die Bereitstellung eines Links zum unmittelbaren Zugriff auf das Dokument.

Dennoch sind die Leistungen einer Suchmaschine nicht denjenigen eines Bibliothekskatalogs vergleichbar. Eine Suchmaschine ist primär für eine Suche nach Fakten konzipiert, für eine Suche nach Dokumenten müsste aber auch sie folgende Anforderungen erfüllen:

**known item search** Suche nach einem bestimmten Dokument (bspw. anhand einer vorliegenden Referenz).

**collocation search** Suche nach zusammengehörenden Dokumenten (bspw. eines Autors, oder Übersetzungen und verschiedene Auflagen eines Werkes).

**subject search** Suche nach Dokumenten, die ein bestimmtes Thema behandeln.<sup>83</sup>

In Katalogen wissenschaftlicher Bibliotheken werden verschiedene Instrumente verwendet, um diese drei Ziele möglichst optimal zu erreichen: Ansetzungsformen, Normdaten/kontrolliertes Vokabular, Verknüpfung verschiedener zusammengehörender Objekte, und grundsätzlich: die Erfassung *strukturierter* Metadaten unter Verwendung eines Da-

---

<sup>82</sup>Zur Generierung universitärer Jahresbibliographien werden Wissenschaftler oftmals auch angehalten, Metadaten für Publikationen zu erfassen, die sie aus urheberrechtlichen Gründen nicht hochladen können.

<sup>83</sup>Vgl. [Eversberg \(2003\)](#), 118f.



tenformats<sup>84</sup>, die verschiedene Indexierungen (nach Verfassern, nach Titeln, nach Erscheinungsjahren, nach Sprachen...) und damit Sucheinstiege ermöglichen.<sup>85</sup>

Suchmaschinen hingegen indexieren das Dokument so, wie es vorliegt, unstrukturiert in einem großen allgemeinen Index. Eine *known item search* ist durch die Eingabe einer längeren Phrase in den meisten Fällen durchführbar<sup>86</sup>, *collocation* und *subject search* jedoch kaum mit befriedigendem Ergebnis realisierbar. Da der von Suchmaschinen genutzte Index auch alle anderen Objekte umfasst, die im Internet verfügbar sind, werden wissenschaftliche Dokumente zudem u. U. in der Ergebnisliste erst an unterer Stelle gelistet.<sup>87</sup>

Dediziert wissenschaftliche Suchmaschinen wie *Google Scholar*<sup>88</sup> entschärfen zumindest das Rankingproblem, Suchergebnisse werden zudem mit wissenschaftlicher Zitierweise ähnlichen und als BibTex<sup>89</sup> herunterladbaren Metadaten präsentiert, die jedoch nicht immer korrekt generiert werden.<sup>90</sup> Auch hier findet allerdings keine eindeutige Autoridentifikation statt, eine Suche nach Schlagworten ist nicht möglich, eine Suche nach Erscheinungsjahren funktioniert nur eingeschränkt.<sup>91</sup>

Die Ziele, deren Erreichung für die Generierung der beschriebenen Funktionalitäten für Repositoriums-inhalte notwendig ist – eindeutige Identifizierung eines Verfassers, einer Publikation, einer Quelle, eines behandelten Themas etc. –, werden somit auch durch Suchmaschinentechnologien bisher nicht erfüllt. Repositorien müssen daher eigenständig qualitativ hochwertige, strukturierte Metadaten zur Verfügung stellen. Da die Strategien der Literatursuche, die Wissenschaftler anwenden, neben speziellen Fachdatenbanken oder Linklisten immer mehr übergreifende, universell verfügbare Instrumente berücksichtigen<sup>92</sup> (vor allem, um schnellen Zugang zu frei verfügbaren Dokumenten zu erhalten<sup>93</sup>), müssen diese Metadaten zudem auf vielfältige Weise nutzbar und einbindbar sein: für

---

<sup>84</sup>Vgl. Eversberg (2003), 122.

<sup>85</sup>Zur Funktionsweise eines Bibliothekskatalogs vgl. z. B. grundsätzlich Gantert u. Hacker (2008), 165ff.

<sup>86</sup>Allerdings ist keine sichere Verifizierung auf *Nichtvorhandensein* eines Dokuments möglich (vgl. Eversberg (2003), 126).

<sup>87</sup>Pieper u. Wolf (2009), 357.

<sup>88</sup>Homepage: <<http://scholar.google.de>> (Abruf: 10.08.2011).

<sup>89</sup>BibTex ist das Datenformat, das im Allgemeinen im vor allem im naturwissenschaftlichen Bereich verbreiteten Satzprogramm LaTeX für die Erzeugung von Literaturangaben und -verzeichnissen genutzt wird.

<sup>90</sup>So wird teilweise der Zeitschriftentitel zum Autor etc.

<sup>91</sup>Vgl. Pieper u. Wolf (2009), 357. *Google Scholar* richtet allerdings seit Juli 2011 einen neuen Service ein, der Publikationen eindeutig einem Verfasser zuordnen und auch Zitationsraten errechnen soll (vgl. Connor (2011).)

<sup>92</sup>Pieper u. Wolf (2009), 356. Die Ergebnisse von Fry u. a. (2009), 90, widersprechen dem allerdings leicht (s. a. S. 14 dieser Arbeit).

<sup>93</sup>Fry u. a. (2009), 60f.

Suchmaschinen wie *Google/Google Scholar*; für OAI-Service-Provider wie *OAIster*<sup>94</sup> und *BASE*<sup>95</sup>; für Bibliotheks-Portale, um auch hier direkt einen Link zum Volltext zur Verfügung stellen zu können; für APIs, die neue Services entwickeln wollen; für virtuelle Forschungsumgebungen; für das Semantic Web. Die Bedeutung der Suchoberfläche des einzelnen IRs (des Ursprungs eines Metadatensatzes) verliert hingegen an Bedeutung, da Forschung selten institutionsbezogen, sondern disziplinär organisiert ist.<sup>96</sup>

## 3.2 Derzeitige Situation der Metadatenerschließung in Institutionellen Repositorien

In IR werden Metadaten gewöhnlich vom Submitter beim Hochladen eines Dokumentes erfasst. Dies kann der bzw. einer der Verfasser des Artikels selbst sein, aber auch eine von ihm beauftragte Person, z. B. eine studentische Hilfskraft oder ein Lehrstuhlmitarbeiter.<sup>97</sup> Eine Endredaktion durch die das Repositoryum betreibende Institution findet in vielen Fällen nicht statt.<sup>98</sup>

Die Eingabe erfolgt durch ein von der Repositoryumssoftware bereitgestelltes Formular, das je nach zu erfassendem Publikationstyp entsprechende Felder anbietet, von denen einige als Pflichtfelder deklariert werden. Informationen zu den zu erfassenden Feldinhalten gibt es entweder kontextsensitiv oder in allgemeinen Hilfetexten (s. Abb. 1<sup>99</sup>).

Eine wichtige Konstante der Dokumentenerschließung ist die für Repositorien entwickelte Schnittstelle Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), über die Service Provider und Aggregatoren auf die Metadatensätze der Repositorien zugreifen. Mindeststandard für diesen Zugriff ist das Datenformat Dublin

<sup>94</sup>Homepage: <<http://www.oclc.org/oaister/>> (Abruf: 10.08.2011).

<sup>95</sup>Homepage: <<http://www.base-search.net/>> (Abruf: 10.08.2011).

<sup>96</sup>Allerdings können Repositorien hier institutionsbezogene Funktionen wie Erzeugung einer universitären Jahresbibliographie oder Bereitstellung von Dokumenten für die Studierenden eines Seminars erfüllen.

<sup>97</sup>In einigen Repositorien gibt es Fremddaten, z. B. in *Goescholar*, dem IR der Universität Göttingen (s. *Goescholar* (2011), Abschnitt 'Externe Daten aus Verlagsvereinbarungen').

<sup>98</sup>Chapman u. a. (2009), 310. Ausnahme ist z. B. das Zurich Open Repository and Archive, das in den Leitlinien eine Kontrolle der Metadaten vorsieht (vgl. *Univ. Zürich* (2011b), Punkt 4). Dies liegt u. U. auch darin begründet, dass die Einträge von ZORA die Grundlage der Akademischen Berichte der Universität Zürich sind (vgl. ebd., Punkt 5).

<sup>99</sup>URL des Eingabeformulars von *BiPrints*: <<http://repositories.ub.uni-bielefeld.de/biprints/uni/?la=de>> (mittlerweile nicht mehr abrufbar, da die Software von BiPrints umgestellt wird (10.08.2011); es handelt sich um ein typisches Eingabeformular der Repositoryumssoftware *OPUS* (vgl. *OPUS* (2011)). URL des Eingabeformulars von *ZORA*: <<http://www.zora.uzh.ch/cgi/users/home?screen=EPrint::Edit&eprintid=48910&stage=core#t>> (Aufruf nur nach Login möglich); verwendet wird die Software *EPrints*. (vgl. *Mazzoni* (2009), Abschn. 'Softwareevaluation').

| Publication Details   |   |
|-----------------------|---|
| <b>Item Sub-type:</b> | Please specify the type of your item.<br>Original work (original research)  |
| <b>Refereed:</b>      | Yes, the item has been refereed.  |
| <b>Status:</b>        | Please state here whether your deposit has been published in final form (with all bibliographic data such as volume, issue and page numbers) or has been published online ahead of print/final form (at this stage without such data).<br>Published in final form |
| <b>Language:</b>      | <div> English<br/> German<br/> French<br/> Italian<br/> Spanish<br/> Abkhazian<br/> Afar<br/> Afrikaans<br/> Akan<br/> Albanian </div>  |
| <b>Journal Title:</b> | The title of the scientific journal. Can be abbreviated with full stops [...] or entered in full.<br>Example: <b>Journal of Experimental Marine Biology and Ecology</b><br>Example: <b>J. Cell Biol.</b>  |
| <b>ISSN:</b>          |   |
| <b>Publisher:</b>     |   |
| <b>Volume:</b>        |   |
| <b>Number:</b>        |   |

(a) Eingabeformular ZORA (Ausschnitt)

#### Anmelden von Postprints für BiPrints

Schritt (1) - (2) - (3)

Bitte beschreiben Sie hier das Dokument, das Sie in BiPrints einbringen möchten.  
Die mit ♦ gekennzeichneten Felder sind obligatorisch auszufüllen (Pflichtelemente).

♦ Dokumentart Zeitschriftenaufsatz

♦ Titel

Untertitel

♦ Verfasser (Nachname, Vorname)

♦ Beteiligte Einrichtung 1

bitte auswählen

Schlagwörter

Bitte trennen Sie die einzelnen Begriffe (max. 5) durch Leerzeichen Komma Leerzeichen (Schlagwort1 , Schlagwort2)

♦ DDC-Sachgruppe

bitte auswählen

Abstract (max. 5000 Zeichen)

(b) Eingabeformular BiPrints (Ausschnitt)

*Abbildung 1 – Beispiele Eingabeformular*

Core (DC) Simple<sup>100</sup> mit seinen 15 Elementen, auch andere – umfangreichere – Datenformate können allerdings angeboten werden, zudem kann das Internformat auch ein proprietäres sein.<sup>101</sup>

Standards für Repositorien auch auf Metadatenebene werden zudem von verschiedenen nationalen und internationalen Initiativen und Projekten ausgearbeitet. Auf deutscher Ebene hat hier besonders die *Deutsche Initiative für Netzwerkinformation* (DINI) Bedeutung, die Repositorien ein Zertifikat verleiht, wenn Mindestanforderungen in Bereichen wie „Sichtbarkeit“, „Unterstützung für Autoren und Herausgeber“ oder „Langzeitverfügbarkeit“ etc. erfüllt werden. Im Kapitel „Erschließung und Schnittstellen“ der entsprechenden Anforderungsliste werden fünf Mindestanforderungen und zusätzlich drei Empfehlungen im Bereich Metadaten gelistet, die sich primär auf die inhaltliche Erschließung beziehen. Die Version 2010 des Zertifikats verfügt zudem über einen zusätzlichen Anhang, der spezielle Festlegungen für die OAI-PMH-Schnittstelle trifft.<sup>102</sup>

Das Projekt der Europäischen Kommission *Digital Repositories Infrastructure Vision for European Research* (DRIVER) hat zum Ziel, über nationale Grenzen hinweg Unterstützung zu leisten für den Aufbau digitaler Repositorien in Europa, sowohl im Bereich Publikationen als auch Forschungsdaten.<sup>103</sup> Neben dem Aufbau eines Suchportals über mehr als 70 europäische Repositorien<sup>104</sup> wurden auch *Guidelines* für den Umgang mit der OAI-PMH-Schnittstelle erarbeitet. Auch diese verstehen sich eindeutig nicht als Katalogisierungsanleitung<sup>105</sup>, geben allerdings umfangreichere Hinweise als das *DINI-Zertifikat* darauf, wie die 15 Elemente des DC-Simple-Formats zu bedienen sind. (Näheres zu den DINI- und DRIVER-Anforderungen s. Kap. 5.)

Es lässt sich daraus schließen, dass Metadaten in Repositorien primär unter dem Aspekt OAI-PMH und dessen Mindeststandard DC Simple behandelt werden, der der Bezugspunkt sowohl der Formulargestaltung zur Dateneingabe wie auch der Empfehlungen von DINI und DRIVER ist. Eine hohe Qualität der Metadaten ist aber auch unabhängig von den Anforderungen des Datenformats – und darüber hinausgehend – notwendig, denn Nutzer, die den hohen Erschließungsstandard anderer Literaturnach-

---

<sup>100</sup>Dublin Core ist ein Datenformat zur Beschreibung von Objekten im Internet. In seiner Kerndefinition *Dublin Core Metadata Element Set*, auch *Dublin Core Simple* genannt, enthält es 15 Elemente, die durch *DCMI Metadata Terms* zum sog. *Qualified Dublin Core* erweitert werden können (vgl. Taylor u. Joudrey (2009), 213ff. u. Gantert u. Hacker (2008), 176). Die Spezifikation ist unter <<http://dublincore.org/>> erhältlich (Abruf: 10.08.2011).

<sup>101</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 26.

<sup>102</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 26ff. u. 47ff.

<sup>103</sup>DRIVER (2008), Vanderfeesten u. a. (2008), 9ff.

<sup>104</sup>Homepage: <<http://search.driver.research-infrastructures.eu/>> (Abruf: 10.08.2011).

<sup>105</sup>Vanderfeesten u. a. (2008), 13.

weise gewohnt sind, werden die IR sonst nicht akzeptieren.<sup>106</sup> Darüber hinaus sind die Metadaten auch die Grundlage der zu entwickelnden Dienstleistungen, die IR, wie in Kap. 2.2 beschrieben, attraktiver machen sollen.

In den folgenden Kapiteln sollen daher Anforderungen an Metadaten hoher Qualität für Dokumente aus IR diskutiert werden.

---

<sup>106</sup>Vgl. Joint (2006), 82.

## 4 Qualität von Metadaten

### 4.1 Definition

Die Katalogisierung der in einer Bibliothek vorhandenen Medien und die Qualität ihres Katalogs gehören seit jeher zu den Kernpunkten in der bibliothekarischen Fachdiskussion. Der Katalog entscheidet über die Nutzbarkeit des Bestandes: Ist ein Medium dort formal und inhaltlich nicht oder nicht korrekt erschlossen, wird es auch nicht genutzt.<sup>107</sup> Auch übergreifende Bibliotheksportale oder sog. Discovery Services, die frei verfügbare elektronische Dokumente in ihre Suchergebnisse mit einbeziehen, benötigen korrekte Metadaten. Umfangreiche Regelwerke wie *Regeln für die alphabetische Katalogisierung in wissenschaftlichen Bibliotheken (RAK-WB)* oder *Anglo-American Cataloging Rules, 2. ed. (AACR2)* sowie auf Verbund- und Bibliotheksebene vorhandene Einzelregelungen bestimmen daher so exakt wie möglich den Umgang mit den einzelnen Elementen der Titelaufnahme. Sie versuchen, eine Standardisierung herzustellen, die für die Nutzbarkeit des Katalogs und den Fremddatenaustausch zwischen den Bibliotheken unerlässlich ist. Ihr Ziel ist die Erfüllung der vier grundlegenden Aufgaben von Metadaten, die 1997 von der *International Federation of Library Associations and Institutions (IFLA) Study Group on the Functional Requirements for Bibliographic Records* folgendermaßen definiert wurden:

- using the data to *find* materials that correspond to the user's stated search criteria
- using the data retrieved to *identify* an entity
- using the data to *select* an entity that is appropriate to the user's needs
- using the data in order to acquire or *obtain* access to the entity described<sup>108</sup>

Die Generierung von Metadaten, die diese grundlegenden Aufgaben erfüllen können, bewegt sich heute in einem Spannungsfeld von Qualität auf der einen und Effizienz auf der anderen Seite, hervorgerufen durch Personalabbau, immer höhere Bestandszuwächse im elektronischen Bereich und neue Anforderungen technischer Art (wie in Kap. 2.2 dieser Arbeit beschrieben), die auch durch nicht-bibliothekarische Communities ins Spiel gebracht werden.<sup>109</sup>

<sup>107</sup>Vgl. Haller (1998), 18 und Haller u. Fabian (2004), 222.

<sup>108</sup>IFLA Study Group on the Functional Requirements for Bibliographic Records (2008), 8.

<sup>109</sup>Bruce u. Hillmann (2004) beschreiben diese Situation (239f.) und appellieren dabei an die verschiedenen Communities zu kooperieren: „Isolation and a tendency to manufacture special solutions for what are really general problems create barriers for coordinated thinking about quality.“

Zwischen Metadaten in Bibliothekskatalogen und in institutionellen Repositorien gibt es allerdings elementare Unterschiede:

1. Metadaten für Bibliothekskataloge werden traditionell von ausgebildeten Bibliothekarinnen und Bibliothekaren, also von Metadaten-Spezialisten, erfasst – in Repositorien hingegen ist ein Submitter ohne fachlichen Hintergrund verantwortlich für die Beschreibung des von ihm hochgeladenen Dokuments.<sup>110</sup> Es ist daher zu untersuchen, welche Qualitätsstandards auf Repositoriums-Metadaten überhaupt angewendet werden können bzw. wie durch technische Unterstützung dennoch ein hoher Qualitätsstandard erreicht werden kann.
2. Es ist nicht ausreichend, wenn Metadaten in ihrer eigenen Umgebung konsistent und verständlich interpretierbar sind – sie müssen diesen Anspruch ebenfalls in aggregierten Sammlungen im Zusammenhang mit Datensätzen aus anderen Repositorien mit unterschiedlichem Erschließungshintergrund (s. a. Abb. 2 auf S. 37) erfüllen<sup>111</sup>, oder auch in gänzlich anderen technischen Umgebungen, für die sie über Schnittstellen bereitgestellt werden.

Die Möglichkeit des Datenaustauschs und Aggregierens über *OAI-PMH* von Datensätzen aus Repositorien stellt eine wichtige Konstante für den Charakter der Metadaten dar. Die Erfassungsformulare der Repositorien orientieren sich weniger inhaltlich an den verbreiteten bibliothekarischen Regelwerken als technisch am Austauschformat *DC Simple*, das für Datenabfragen und -aggregationen der Mindeststandard ist.<sup>112</sup> Wie bereits beschrieben, gibt es hierfür wiederum diverse Festlegungen und Empfehlungen – aber keine Regelwerke! – von Arbeitsgruppen der Repositoriumsbetreiber, z. B. auf deutscher Ebene durch das *DINI-Zertifikat* oder auf europäischer Ebene durch die *DRIVER Guidelines*.<sup>113</sup>

Die 15 Elemente von *DC Simple* sind kaum mit der – von Repositoriumsnutzern ohnehin nicht zu leistenden – Detailliertheit der Angaben, die bibliothekarische Regelwerke vorsehen, vergleichbar, und somit auch nur schwerlich ihre Qualität. Die Nutzungsmöglichkeiten der institutionellen Repositorien stellen Anforderungen, auf die die Regelwerke nicht eingehen, für die auf der anderen Seite aber auch *DC Simple* nicht ausreichend ist. Qualitätskriterium der Metadaten in *IR* muss daher nicht nur sein, ob bibliothekarische

---

<sup>110</sup>Vgl. Guy u. a. (2004).

<sup>111</sup>Vgl. Guy u. a. (2004).

<sup>112</sup>Lagoze u. Sompel (2008), 3.4.

<sup>113</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011); Vanderfeesten u. a. (2008).

Ansprüche erfüllt sind oder ob ein bestimmtes Datenformat korrekt bedient wird, sondern vor allem, ob die funktionalen Anforderungen des modernen Datenaustauschs und der Wissenschaftskommunikation erfüllt werden können.<sup>114</sup> IR sollen der Verbreitung freier wissenschaftlicher Publikationen dienen – an der Erreichung dieses Ziels muss sich ihre Qualität messen lassen. Dazu gehört ebenfalls die Fähigkeit der Metadaten, über Schnittstellen, Austauschformate oder Linked-Data-Konzepte in anderen, übergreifenden Sammlungen und Anwendungen präsentiert und genutzt werden zu können.

Für die Erzeugung qualitativ hochwertiger Metadaten werden daher benötigt:

- *allgemeine Qualitätskriterien* für Metadaten für elektronische Ressourcen
- *funktionale Anforderungen* an die Metadatenätze und -elemente für elektronische Ressourcen

## 4.2 Allgemeine Qualitätskriterien für Metadaten elektronischer Ressourcen

Allgemeine Qualitätskriterien sollen bestimmen, wie die Eigenschaften eines Metadatenatzes bzw. -elements und ihr Inhalt beschaffen sind, z. B. korrekte Rechtschreibung, standardisierte Erfassung, Vollständigkeit. Wie oben beschrieben, können allgemeine bibliothekarische Katalogisierungsregeln nur bedingt auf Metadaten elektronischer Dokumente übertragen werden, die Einhaltung von Qualitätsstandards ist aber auch hier entscheidend für die Nutzbarkeit der beschriebenen Ressource. Bruce u. Hillmann (2004)<sup>115</sup> konstatieren, dass zwar zahlreiche neue Datenformate und -standards entstanden sind, Qualitätskriterien bei der Begutachtung der zahlreichen Anträge für Digitalisierungsprojekte aber kaum berücksichtigt werden. Unter den Analysen dieser Kriterien, die seit Ende der 90er Jahre durchgeführt wurden, sind drei grundlegende hervorzuheben: Moen u. a. (1997), Bruce u. Hillmann (2004) und Stvilia u. a. (2004)/Shreeves u. a. (2005).

Für den Aufbau eines staatlichen Regierungsinformationssystem evaluierten Moen u. a. (1997) Datensätze nach insgesamt 23 in der informationswissenschaftlichen Fachliteratur identifizierten Qualitätskriterien. Um eine koordinierte Herangehensweise an diese Analyse zu erreichen, teilten sie die Kriterien in insgesamt vier Kategorien ein:

<sup>114</sup>Vgl. Guy u. a. (2004): „[...] *‘high quality metadata supports the functional requirements of the system it is designed to support’*, which can be summarised as *‘quality is about fitness for purpose’*“; Hillmann u. a. (2004), 7: „The utility of metadata can best be evaluated in the context of services provided to end-users.“; vgl. auch Foulonneau u. Riley (2008), 54ff.

<sup>115</sup>Bruce u. Hillmann (2004), 240f.



- *accuracy*: typographische und formale Korrekt- und Genauigkeit
- *completeness*: Vollständigkeit des Elementsets, des einzelnen Elements und des Inhalts
- *profile*: Beschreibung von Datensatz – und Objekttypen, Herkunft (in aggregierten Sammlungen)
- *serviceability*: Suchbarkeit, Konsistenz, Lesbarkeit

Stehen die von [Moen u. a. \(1997\)](#) untersuchten Qualitätskriterien noch in engem Zusammenhang mit den analysierten Metadatensätzen ihres Projekts, versuchen [Bruce u. Hillmann \(2004\)](#) vor allem in Bezug auf Digitalisierungsprojekte, rasch wachsende Sammlungen elektronischer Dokumente und die Anforderungen des Datentauschs, *allgemeine* Kriterien für die Qualität von Metadaten zu entwerfen: abstrakt, um auf verschiedene Szenarien anwendbar zu sein; pragmatisch und nicht dogmatisch, um ökonomischen Zwängen gerecht werden zu können. Sie ziehen hierzu sieben unterschiedliche Dimensionen in Betracht, die in Kap. 4.3 näher beschrieben werden: *completeness*, *accuracy*, *provenance*, *conformance to expectations*, *consistency and coherence*, *timeliness*, und *accessibility*.

[Stvilia u. a. \(2004\)](#) und erneut [Shreeves u. a. \(2005\)](#) gehen methodologisch abweichend vor: Sie konstruieren ihr Modell nicht vom Metadatensatz oder einzelnen Elementen ausgehend, sondern vom Nutzen her, den diese im Informationsprozess erfüllen sollen. Drei übergeordnete Kategorien werden hierzu definiert, die von der Art der Beziehung der einzelnen Elemente untereinander, zu Standards, Nutzungszwecken oder dem beschriebenen Objekt bestimmt werden:

***Intrinsic Information Quality*** umfasst Kriterien, die in Bezug auf das Metadatenobjekt selbst oder auf verwendete Standards und Encodings beurteilt werden können. Beispiele sind: Tippfehler; Konsistenz der Struktur der Datumsangaben; Konsistenz der Struktur der Reihenfolge von Vorname und Nachname; Konsistenz des Sets der verwendeten Elemente; Anwendung des Datenformats; freies/kontrolliertes Vokabular zur inhaltlichen Beschreibung.

***Relational/Contextual Information Quality*** umfasst Kriterien, die in Bezug auf ein Objekt außerhalb des Metadatensatzes messbar sind: Korrekte Angabe des Titels/des Verfassers/der Quelle; community-gemäße inhaltliche Beschreibung des Objekts; für die Verwendung des Metadatensatzes notwendige Vollständigkeit.

**Reputational Information Quality** beschreibt das Maß an Vertrauen, das ein Nutzer in die Informationen eines Metadatensatzes haben kann. Die Reputation eines Metadatensatzes wird erhöht, wenn er aus einer vertrauenswürdigen, dem Nutzer bekannten Quelle stammt oder beispielsweise die inhaltliche Erschließung durch eine der Community gemäßen Klassifikation erfolgt. In aggregierten Sammlungen kann das Vorkommen mehrerer Datensätze, die das gleiche Objekt beschreiben, einerseits das Vertrauen verringern, wenn sie sich in einigen Elementen voneinander unterscheiden, andererseits bei vollständiger Gleichheit auch erhöhen (Bestätigung der Angaben durch doppeltes Vorkommen).<sup>116</sup>

Innerhalb dieser Grundkategorien definieren [Shreeves u. a. \(2005\)](#)<sup>117</sup> wiederum 16 einzelne Kriterien, anhand derer Qualität messbar sein soll. Einige der Kriterien können gleichzeitig verschiedenen Grundkategorien zugeordnet werden, ihre Anwendung erfolgt dann aus dem Blickwinkel der jeweiligen Kategorie: Z. B. wird *accuracy* einerseits in Bezug auf einen allgemeinen Standard oder ein Wörterbuch gemessen (*intrinsic information quality*) und andererseits in Bezug auf das Objekt, das die Metadaten beschreiben (*relational information quality*).

Die drei zitierten Modelle überschneiden sich vielfach in den verwendeten Kriterien, so kommen z. B. bei allen *accuracy* und *completeness* vor. [Shreeves u. a. \(2005\)](#) nehmen daher eine Zuordnung ihrer Kriterien<sup>118</sup> zu denjenigen von [Bruce u. Hillmann \(2004\)](#) vor; deren kompakteres, anwendungsorientierteres Modell soll im Folgenden näher beschrieben und später angewendet werden.

### 4.3 Allgemeine Qualitätskriterien für Metadaten elektronischer Dokumente nach Bruce/Hillmann

[Bruce u. Hillmann \(2004\)](#) versuchen einen pragmatischen, nicht idealistischen Ansatz, Kriterien für Metadaten elektronischer Dokumente zu definieren. Ihre Qualität müssten die Datensätze vor allem in umfangreichen aggregierten Sammlungen beweisen, die Metadaten verschiedenster Standards umfassen. Konzeptionen von Metadaten sollten also nicht zu sehr auf das eigene Projekt bezogen, sondern allgemeiner ausgerichtet sein. Alle theoretisch vorstellbaren Ziele könnten wegen äußerer Zwänge – seien sie finanzieller, personeller oder technischer Natur – ohnehin nicht erreicht werden.<sup>119</sup> Entwickelt

<sup>116</sup>[Stvilia u. a. \(2004\)](#), 114 u. 116f.; [Shreeves u. a. \(2005\)](#), 224f. u. 233.

<sup>117</sup>[Shreeves u. a. \(2005\)](#), 233.

<sup>118</sup>[Shreeves u. a. \(2005\)](#), 224.

<sup>119</sup>Vgl. [Bruce u. Hillmann \(2004\)](#), 242: „Too frequently, implementers fall into the trap of bipolar thinking, making the perfect the enemy of the good.“

werden hier auf der Grundlage des von *Statistics Canada* entworfenen „Quality Assurance Frameworks“<sup>120</sup> sieben Messwerte, durch die sich die Qualität von Metadaten für elektronische Dokumente bestimmen lassen soll:

### **Completeness**

- Das Objekt soll so vollständig wie möglich, aber auch nur so vollständig wie nötig beschrieben werden.
- Das ausgewählte Elementset soll möglichst auf die vollständige Sammlung von Objekten angewandt werden.
- Alle Elemente müssen erfasst sein, die zur eindeutigen Beschreibung des Objektes und zu seiner Auffindbarkeit, auch in externen aggregierten Sammlungen, notwendig sind.<sup>121</sup>

### **Accuracy**

- Elemente sollen mit eindeutigem Inhalt gefüllt werden.
- Tippfehler etc. sollten nicht vorhanden sein.
- Datensätze, die das Kriterium der *Accuracy* nicht erfüllen, sind u. U. nicht mehr auffindbar.<sup>122</sup>

### **Provenance**

- Die Herkunft des Datensatzes sollte klar erkennbar sein.
- Die Veränderungen, die ein Datensatz seit seiner Entstehung erfahren hat, sowie deren Urheber sollten erkennbar sein.
- Der Ersteller des Datensatzes und der an ihm bisher durchgeführten Veränderungen sollte klar erkennbar sein (Submitter/Autor des beschriebenen Dokuments, Redakteur/fachliches Personal, automatisches Verfahren).
- Falls die Erstellung des Datensatzes auf einem Regelwerk basiert (wie z. B. [RAK-WB](#) oder [AACR2](#) ), sollte auch dies erkennbar sein.

---

<sup>120</sup>Bruce u. Hillmann (2004), 242.

<sup>121</sup>Bruce u. Hillmann (2004), 243. Vgl. auch Park u. Tosaka (2010), 699; Shreeves u. a. (2005), 226f. u. 233.

<sup>122</sup>Bruce u. Hillmann (2004), 243. Vgl. auch Park u. Tosaka (2010), 699.

- Diese Angaben sollten möglichst bis auf Elementebene zurückführbar sein.<sup>123</sup>

### ***Conformance to expectations***

- Das Elementset sollte den Erwartungen entsprechen, die die rezipierende Community an es hat. Dies ist sowohl relevant für die Suche, aber auch für den Submitter, der das Elementset füllen soll.
- Das Elementset sollte über keine Elemente verfügen, die für die rezipierende Community irrelevant oder die für den größten Teil der Sammlung gar nicht zu füllen sind.
- Zwischen den hohen Erwartungen der Community und den durch die Datenlieferanten am einfachsten zu generierenden Daten sollte ein realistisches Mittel gefunden werden.<sup>124</sup>

### ***Logical consistency and coherence***

- Die Objekte einer Sammlung sollten möglichst immer auf die gleiche Art und Weise beschrieben werden. Dies gilt sowohl für das Konzept der Inhalte wie auch für die Struktur des Datensatzes und des einzelnen Elements.
- Die konsistente Angabe gleicher Informationen in gleichen Elementen ermöglicht die Generierung konsistenter Indizes für Suchen und Browsen.
- Die Ermöglichung von Referenzierungen zu anderen Systemen erhöht die Konsistenz.
- *Logical consistency and coherence* sind besonders wichtig für die Austauschbarkeit von Datensätzen.<sup>125</sup>

### ***Timeliness: currency and lag***

- (*currency*) Die Beschreibung eines Objektes kann nie als endgültig abgeschlossen gelten, sondern muss aktualisiert werden (können), da sich die äußeren Bedingungen oder das Konzept seiner Beschreibung ändern können.

---

<sup>123</sup>Bruce u. Hillmann (2004), 243f.

<sup>124</sup>Bruce u. Hillmann (2004), 244.

<sup>125</sup>Bruce u. Hillmann (2004), 244f. Vgl. auch Park u. Tosaka (2010), 699; Shreeves u. a. (2005), 227 u. 233.

- (*lag*) Die Beschreibung eines Objektes muss der verbesserten Informationslage über das Objekt angepasst werden.<sup>126</sup>

### **Accessibility**

- Metadaten müssen technisch zugänglich, also von Standardprogrammen interpretierbar und über Standardschnittstellen frei abrufbar sein.
- Metadaten müssen intellektuell zugänglich, d. h. ausreichend dokumentiert sein.<sup>127</sup>

Anwendbar und teilweise automatisch überprüfbar sind diese von Bruce/Hillmann aufgestellten Kriterien von drei Gesichtspunkten aus:

**Struktur** Prüfen auf Vorhandensein bzw. Analyse von Syntax, Datensatz-Wrapper, Namespaces, Identifier, Erstellungsdatum etc.

**Inhalt** Analyse/Prüfen auf Vorhandensein von kontrolliertem Vokabular, ausreichenden fachübergreifenden Informationen, ausreichend community-gemäßen Informationen etc.

**Dokumentation** Application Profile, Angaben zur Herkunft der Datensätze und zu Änderungen etc.<sup>128</sup>

## **4.4 Funktionale Anforderungen**

Funktionale Anforderungen legen fest, welchen definierten *Zweck* ein Element bzw. ein Metadatensatz erfüllen soll. (Im Gegensatz dazu werden *Eigenschaften* von Elementen, wie z. B. das Datenformat oder ein Encoding Typ, von nicht-funktionalen Anforderungen definiert.)<sup>129</sup> Voraussetzung für die Erstellung funktionaler Anforderungen ist eine Anforderungsanalyse. Als eine entsprechende Grundlage der in Kap. 5 erarbeiteten Punkte werden in dieser Arbeit daher die in Kap. 2.2 beschriebenen Ansprüche der modernen Wissenschaftskommunikation genutzt.

In der Literatur existieren zahlreiche Beschreibungen von verschiedenen Digitalisierungsprojekten und angestrebten Funktionalitäten, die die Beschreibung der digitalisierten Objekte leisten sollen.<sup>130</sup> Notwendig ist allerdings auch eine allgemeinere, theo-

<sup>126</sup>Bruce u. Hillmann (2004), 245ff.

<sup>127</sup>Bruce u. Hillmann (2004), 247f.

<sup>128</sup>Bruce u. Hillmann (2004), 249ff.

<sup>129</sup>Vgl. Robertson u. Robertson (2006), 9f: „Functional requirements are things the product must do.“ bzw. „Nonfunctional requirements are qualities the product must have.“

<sup>130</sup>So beziehen sich beispielsweise sieben von zehn Artikeln des Themenheftes „Metadata and Open Access repositories“ der Zeitschrift *Cataloging & Classification Quarterly* (47 (2009), 3/4) auf einzelne Repositorien.

retischere Herangehensweise, denn die gewünschten Funktionalitäten sollten möglichst umgebungsunabhängig erfüllbar sein.

Aus diesem Grunde müssen nicht nur interne funktionale Anforderungen für das eigene Repositorium festgelegt werden, sondern zusätzlich externe<sup>131</sup>: Jede einzelne Ebene, von der umfangreichen aggregierten Sammlung bis zum möglichen Teilinhalt eines Elements, sollte bei der Konzeption der Metadatenerschließung berücksichtigt werden. In Abb. 2 wird die Komplexität der möglichen Kombinationen illustriert: Jedes einzelne Element muss auf jeder möglichen Ebene über eine verständliche und interpretierbare Bedeutung verfügen.

Es darf dabei auch hier nicht zur dominierenden Einschränkung werden, dass das Standardaustauschformat der Repositorien, DC Simple, nur 15 Elemente umfasst und keine Qualifier vorsieht. Datensätze können auch über das wesentlich umfangreichere Metadata Object Description Schema (MODS)<sup>132</sup>, über *Qualified DC* oder in Deutschland über *XMetaDissPlus*<sup>133</sup> ausgetauscht werden. Metadaten können auch als Linked-Open-Data unter Nutzung bibliothekarischer und anderer Ontologien in Resource Description Framework (RDF)<sup>134</sup> veröffentlicht und damit Teil des Semantic Web werden. Funktionale Anforderungen sollten daher unabhängig von zur Zeit verfügbaren Datenformaten entworfen werden.

---

<sup>131</sup>Guy u. a. (2004).

<sup>132</sup>MODS ist ein von der Library of Congress (LoC) entwickeltes bibliothekarisches Datenformat im XML-Format, das eine enge Verwandtschaft mit MARC 21 hat, aber nicht dessen Komplexität aufweist (vgl. Foulonneau u. André (2008), 29 u. Taylor u. Joudrey (2009), 219ff.). Die Spezifikation ist unter <<http://www.loc.gov/standards/mods/>> vorhanden (Abruf: 10.08.2011).

<sup>133</sup>*XMetaDissPlus* ist ein speziell für Hochschulschriften entwickeltes Datenformat für die Ablieferung an die Deutsche Nationalbibliothek (DNB). Vgl. Deutsche Nationalbibliothek (2010), hier ist auch die Spezifikation enthalten.

<sup>134</sup>RDF ist ein standardisiertes Daten(austausch)format zur Beschreibung und Verknüpfung von Webressourcen, die semantisch als sog. „Tripel“ (Prädikat – Subjekt – Objekt) realisiert werden (vgl. Taylor u. Joudrey (2009), 109ff.). Die Spezifikation ist unter <<http://www.w3.org/RDF/>> vorhanden.

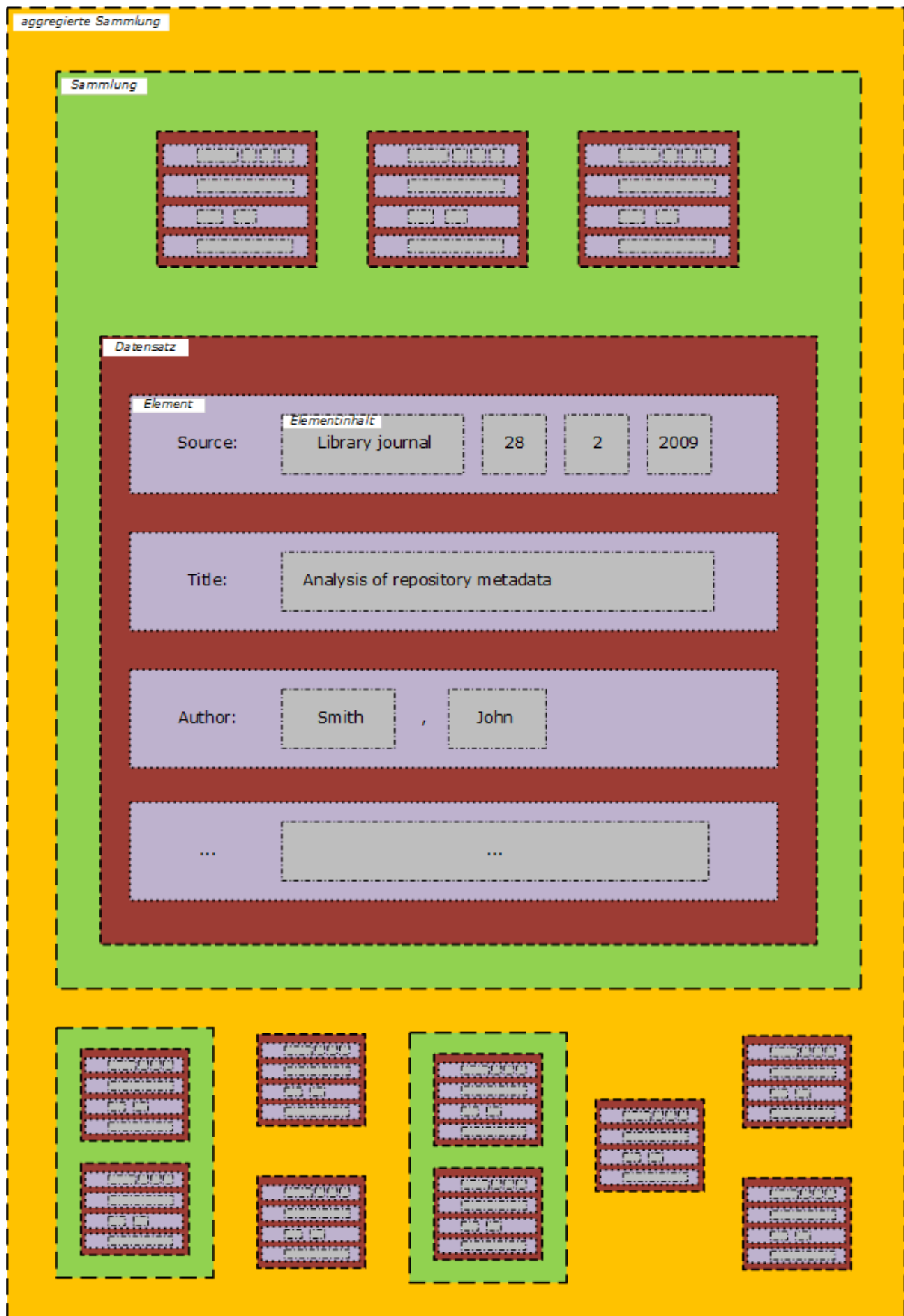


Abbildung 2 – Ebenen von Metadatensätzen aus Repositorien

## 5 Qualitätskriterien und funktionale Anforderungen an Metadaten für selbstarchivierte Zeitschriftenpublikationen in Institutionellen Repositorien

Wie in den vorangehenden Kapiteln dargelegt, besitzt die Qualität der Erschließung selbstarchivierter Dokumente eine hohe Relevanz für die Sichtbarkeit der hochgeladenen Dokumente im Web und für ihre Integration in andere Anwendungen, z. B. für die Erzeugung von dynamisch erzeugten Literaturlisten auf Homepages, Zitationsanalysen, Verlinkungen im Semantic Web, aber auch für kollaboratives, verteiltes Arbeiten mehrerer Wissenschaftler an einer gemeinsamen Publikation.

Um die diese Qualität bestimmenden Faktoren definieren zu können, sollen im Folgenden beispielhaft die einzelnen Elemente, die zur Beschreibung eines in einer Peer-Review-Zeitschrift erschienen Artikels notwendig sind, näher analysiert und Anforderungen für ihre Anwendung erstellt werden. Der Publikationstyp *Zeitschriftenartikel* wurde gewählt, weil er besonders im Fokus der Open-Access-Bemühungen um Selbstarchivierung steht.

Beispieldatensätze aus verschiedenen deutschen und schweizerischen [IR](#) in unterschiedlichen Datenformaten sind im Anhang [A.1](#) bis [A.5](#) zu finden. Hieraus wurde folgende zu untersuchende Elementinhalte zusammengefasst:

- Titel
- Verfasser
- sachliche/inhaltliche Erschließung
- Quelle
- Affiliation
- Eigenschaften der Publikation (Identifier, Sprache, Publikationstyp, Begutachtungstatus, Version, temporäre Zugriffsbeschränkungen)

Es sollen hier nur die von den Ablieferern selbst erfassten Angaben behandelt werden. Es handelt sich dabei vorwiegend um sog. „deskriptive“ Metadatenelemente, die zum Erfüllen der von der [IFLA](#) definierten Aufgaben notwendig sind (vgl. S. [28](#)), weniger



um technisch-administrative, wie sie beispielsweise für Zwecke der Langzeitarchivierung benötigt werden.<sup>135</sup>

Nach einer kurzen Beschreibung des jeweiligen Elements dienen als Grundlage der Analyse die dediziert metadatenbezogenen Mindestanforderungen und Empfehlungen des *DINI-Zertifikats* und die *DRIVER Guidelines*. Da diese im engen Zusammenhang mit dem OAI-PMH-Austauschformat DC Simple stehen, werden zusätzlich auch die grundlegende DC-Elementbeschreibung (Hillmann (2005)) und das darauf basierende *Scholarly Works Application Profile (SWAP)*<sup>136</sup> mit einbezogen. Anschließend werden für jedes Element anhand der erläuterten allgemeinen Kriterien von Bruce u. Hillmann (2004) (s. Kap. 4.3) Qualitätsansprüche ausgearbeitet. Unter Berücksichtigung der in Kap. 2.2 dargestellten aktuellen Entwicklungen in der Wissenschaftskommunikation folgt dann ein Entwurf von elementspezifischen, datenformatunabhängigen funktionalen Anforderungen. Zum Schluss werden aus den Ergebnissen generelle Schlüsse gezogen.

## 5.1 Elementübergreifende Kriterien

### 5.1.1 Vorgaben DINI/DRIVER

- In den im Format Dublin Core ausgelieferten Datensätzen sind die Elemente *title*, *creator*, *date*, *type* und *identifier* Pflicht.<sup>137</sup> DRIVER nennt zusätzlich noch die Elemente *subject* und *description* als obligatorisch, wenn sie anwendbar sind.<sup>138</sup>
- Die für die Erschließung der Dokumente verwendeten Regelungen wurden explizit formuliert und werden für Nutzer des Publikationsservices (Autoren bzw. Herausgeber und Leser) online bereitgestellt.<sup>139</sup>

### 5.1.2 Allgemeine Qualitätskriterien

- *accessibility*: Verwendung von offenen, standardisierten Datenformaten für den Austausch oder die Präsentation als Linked Data; Vorhandensein von Informationen über die verwendeten Elementsets und ihre Inhalte.
- *completeness*: Vollständigkeit des auf S. 38 definierten Angaben.

---

<sup>135</sup>Zum Unterschied zwischen deskriptiven und administrativen Metadaten vgl. Taylor u. Joudrey (2009), 96ff.

<sup>136</sup>SWAP (2009)

<sup>137</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 62 (M.A.3-1).

<sup>138</sup>Vanderfeesten u. a. (2008), 57f.

<sup>139</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 27 (M.6-1).

- *timeliness*: Die Verwendung von URIs für einzelne Elemente erleichtern Anpassungen, da Veränderungen ausschließlich am referenzierten (Norm-)Datensatz vorgenommen werden müssen.
- *provenance*: Die Kenntnis über ein Regelwerk kann zu Hinweisen über den spezifischen Gebrauch von Elementen und ihrem Inhalt führen und somit Unklarheiten vermeiden helfen.
- *provenance*: Ein Datensatz kann u. U. mehrere Veränderungen/Erweiterungen zur Verbesserung der Qualität erfahren<sup>140</sup>, deren Nachvollziehbarkeit und Wertigkeit relevant sein kann für die Interpretation des Datensatzes sowie weitere Modifikationen.

### 5.1.3 Funktionale Anforderungen

#### **Anforderung Ü-1: Die Originalherkunft des Metadatensatzes muss klar erkennbar und referenzierbar sein.<sup>141</sup>**

In aggregierten Sammlungen oder anderen Anwendungsumgebungen werden Metadatensätze aus verschiedenen Repositorien zusammengebracht und unter Umständen erweitert/verbessert. Daher sollte es möglich sein, einen bestimmten Metadatensatz über einen Identifikator ansprechen zu können, um ihm in seinen verschiedenen Vorkommen bzw. zumindest dem Ursprungssatz diese Änderungen anbieten zu können.

## 5.2 Titel

Der Titel ist der Name des beschriebenen Objekts.<sup>142</sup> Im Gegensatz zu allgemeinen Webressourcen, bei denen seine Bestimmung oft schwer fällt, ist der Titel bei einem Zeitschriftenartikel eindeutig und meist vom Wissenschaftler, der das Objekt hochlädt, ursprünglich selbst festgelegt worden.

<sup>140</sup>Ein solches Verfahren wird u. a. von [Hillmann u. a. \(2004\)](#) ausführlich beschrieben.

<sup>141</sup>Zur leichten Identifizierung werden die einzelnen funktionalen Anforderungen, entsprechend der Liste der zu untersuchenden Angaben auf S. 38 wie folgt bezeichnet, jeweils mit fortlaufender Zählung: Ü = übergreifende Anforderungen, T = Titel, V = Verfasser, S = sachliche/inhaltliche Erschließung, Q = Quelle, A = Affiliation, P = bezogen auf die Publikationseigenschaften.

<sup>142</sup>[Vanderfeesten u. a. \(2008\)](#), 59; [Hillmann \(2005\)](#), Abschn. 4.1.

### 5.2.1 Vorgaben DC Simple/SWAP

- Wortwahl und -reihenfolge sowie Orthographie vorlagegemäß übernehmen, Groß-/Kleinschreibung nur wie im normalen Sprachgebrauch, Interpunktion kann abweichen.<sup>143</sup>

### 5.2.2 Vorgaben DINI/DRIVER

- Wortwahl und -reihenfolge sowie Orthographie vorlagegemäß übernehmen, Groß-/Kleinschreibung nur wie im normalen Sprachgebrauch.<sup>144</sup>
- Untertitel im selben Element mit „:“ an den Titel anhängen.<sup>145</sup>

### 5.2.3 Allgemeine Qualitätskriterien

- *accuracy*: Muss vollständig (inkl. Untertitel) und darf nicht gekürzt sein, auch einzelne Wörter nicht abkürzen.

### 5.2.4 Funktionale Anforderungen

**Anforderung T-1: Ein Titel soll immer auch in einer englischen Form auffindbar sein.**

Um die Sichtbarkeit einer Publikation zu erhöhen, sollte ein nicht-englischer Titel immer auch in einer englischen Version suchbar sein (vgl. auch [Anforderung S-4](#)). Besonders relevant ist dies für Originaltitel in nichtlateinischen Schriften wie Japanisch, Arabisch, Hebräisch etc., da diese international in Publikationen und Schriftenverzeichnissen oft nur mit der englischen Titelübersetzung zitiert werden.

## 5.3 Verfasser

Der oder die Verfasser sind der oder die Verantwortliche(n) für die Erstellung des Inhalts eines Dokuments.<sup>146</sup> Im Falle der hier behandelten Publikationsform Zeitschriftenartikel handelt es sich dabei immer um Personen. Institutionen, die mit der Publikation im Zusammenhang stehen, werden im Kap. 5.6 behandelt, als Projektfinanzierer auch in Kap. 5.7.

---

<sup>143</sup>SWAP (2009).

<sup>144</sup>Vanderfeesten u. a. (2008), 59.

<sup>145</sup>Vanderfeesten u. a. (2008), 59.

<sup>146</sup>Vanderfeesten u. a. (2008), 59.

Die Erfassung von Personennamen dient im Idealfall der Auffindbarkeit von Publikationen einer Person – in der Mehrheit sind allerdings Publikationen nur Personennamen zuzuordnen (s. [Anforderung V-1](#)). Die Identifizierung von individuellen Personen kann der Generierung von Literaturlisten dienen oder für Statistiken darüber genutzt werden, wie oft ein Verfasser in einer bestimmten Zeitschrift (beispielsweise mit hohem Impact Factor) veröffentlicht hat.

Bei Personennamen ist eine standardisierte Erfassung besonders relevant, um identische Namen abgleichen zu können.

### 5.3.1 Vorgaben DC Simple/SWAP

- Mehrere Personen in getrennten Elementen aufführen; die Nennung soll in der Form „Nachname, Vorname“ erfolgen.<sup>147</sup>
- Für die Nennung eines Namens immer die gleiche Form verwenden, auch wenn er in verschiedenen Vorlagen abweichend erfasst ist.<sup>148</sup>

### 5.3.2 Vorgaben DINI/DRIVER

- Für das Element *creator* folgende Binnenstruktur verwenden: „Nachname, Vorname“.<sup>149</sup> DRIVER erwähnt zusätzlich Präfixe und Namenszusätze und verweist auf den „APA bibliographic writing style as in a reference list“ als entsprechendes *encoding scheme*.<sup>150</sup>
- DRIVER fordert die Verwendung eines eindeutigen Identifiers für jeden Autor, vorzugsweise einen International Standard Name Identifier (ISNI). Die globale Eindeutigkeit (mittels eines eigenen Namespaces) sowie die Persistenz eines eigenen Identifiers sowie seine Auflösung muss durch das Repositorium gewährleistet werden.<sup>151</sup>

### 5.3.3 Allgemeine Qualitätskriterien

- *completeness*: Verfasserangabe muss enthalten: Nachname, Vorname und weitere Namensbestandteile, falls vorhanden und notwendig, um eine Person zu identifizieren.

---

<sup>147</sup>Hillmann (2005), 4.8.; [SWAP \(2009\)](#).

<sup>148</sup>[SWAP \(2009\)](#).

<sup>149</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 63 (M.A.3-4).

<sup>150</sup>Vanderfeesten u. a. (2008), 59ff. u. 121.

<sup>151</sup>Vanderfeesten u. a. (2008), 112f.

- *accuracy*: Vornamen auflösen, um die Möglichkeit von Namensdoppelungen zu minimieren.
- *consistency*: Verfasser immer in der gleichen Art und Weise angeben (z. B. „Nachname, Vorname“), ebenfalls müssen Präfixe sowie weitere Namensbestandteile immer einheitlich erfasst sein.
- *consistency*: Die Verwendung von URIs ermöglicht eine Referenzierung zu Normdateien und Thesauri und damit zu normierten, standardisierten und konsistenten Angaben.
- *timeliness*: Eine Anpassung ist erforderlich, wenn der Name des Verfassers sich ändert (z. B. durch Heirat).

### 5.3.4 Funktionale Anforderungen

#### **Anforderung V-1: Das Metadatenelement für den Autor der Publikation soll eine individuelle Personen repräsentieren, nicht nur einen Namen.**

Die Erfassung von Personennamen beim Hochladen elektronischer Dokumente führt in den untersuchten Repositorien nicht zur Identifizierung einer individuellen Person, sondern nur zum Abspeichern eines einfachen Strings.<sup>152</sup> Dies hat zur Folge, dass verschiedene Autoren gleichen Namens nicht unterscheidbar sind und in Katalogen und Suchmaschinen durch zufällige Namensgleichheit inhaltlich nicht zusammenhängende Publikationen miteinander in Verbindung gebracht oder auch falschen Personen Forschungsergebnisse zugeschrieben werden können.

Im Prinzip stellen sich bei der Erfassung von Personen in Repositorien die gleichen Herausforderungen wie in konventionellen Bibliothekskatalogen: Verschiedene Schreibweisen von Namen, aufgelöste/nicht aufgelöste Vornamen, Namenswechsel nach Heirat etc.<sup>153</sup> In aggregierten Sammlungen gewinnen diese Probleme für eine eindeutige Personenidentifizierung noch an Relevanz. Warner (2010) zeigt für das fachliche Repository *arXiv* auf, dass zahlreiche Accounts verschiedener User auf den selben Namen lauten. So gibt es 100 Accounts für den Namen „Yang, Z“ und 97 für „Lee, J“.<sup>154</sup> Dies macht deutlich, dass Personen, selbst wenn sie standardisiert erfasst werden, nicht ausschließlich anhand ihres Namens unterschieden werden können.

<sup>152</sup>In den Beispielsätzen A.1 – A.5 werden als Verfasser nur reine Namensstrings angegeben, keine individualisierenden Angaben.

<sup>153</sup>Vgl. Salo (2009), 250f.

<sup>154</sup>Warner (2010), 2.

Um aus Metadaten der Repositorien aber personenbezogene Services zu generieren, müssen die dort hochgeladenen Publikationen Individuen – und nicht nur Namen – zugeordnet werden können. Wissenschaftler könnten dann Literaturlisten erzeugen und auf ihrer Homepage dynamisch verlinken; in Katalog- und Suchmaschinenergebnissen könnten unterschiedliche Verfasser gleichen Namens eindeutig selektiert werden (drill down); Alerting-Dienste oder RSS-Feeds könnten pro Person und nicht nur pro Namen angeboten werden.

Um Namensgleichheiten zu vermeiden, sollte der Vorname aufgelöst werden.<sup>155</sup> Für eine Individualisierung ist bei der Erschließung die Erfassung weiterer Informationen notwendig, z. B. das Geburtsjahr. Aus diesen Angaben zur Person kann ein wiederverwendbarer lokaler Normdatensatz erstellt werden, der anhand eines eindeutigen Identifiers zur Verknüpfung in allen Publikationen eines Autors dient, vergleichbar einer Normdatenverknüpfung in konventionellen Bibliothekskatalogen. Dieser Identifier kann bei wiederholten Submit-Vorgängen leicht wiederverwendet und auch einem Useraccount des Repositoriums zugeordnet werden. Ein Problem hierbei ist allerdings, dass Zeitschriftenpublikationen oft von mehreren Autoren verschiedener Institutionen verfasst wurden und dem einzelnen Submitter u. U. nicht für alle weiteren Beteiligten persönliche Angaben bekannt sind.<sup>156</sup>

#### **Anforderung V-2: Die Person soll auch über die Grenzen des eigenen Repositoriums hinaus eindeutig identifizierbar sein.**

Ein unter [Anforderung V-1](#) beschriebener Identifier, der nur im Rahmen seines eigenen Repositoriums Gültigkeit hat, ermöglicht noch nicht die Zuordnung von Publikationen zu individuellen Autoren in aggregierten Sammlungen, Metasuchmaschinen oder Bibliothekskatalogen.<sup>157</sup> Wissenschaftler wechseln während ihrer Laufbahn öfters ihre Institution, daher können sich ihre Publikationen auf verschiedene institutionelle, aber auch fachliche Repositorien verteilen.<sup>158</sup> Die Eindeutigkeit einer Person innerhalb *eines* Repositoriums ist daher nicht ausreichend. Die Referenzierung der Person auf einen Da-

---

<sup>155</sup>In einigen Repositorien, wie z. B. [ZORA](#), sollen Vornamen allerdings grundsätzlich abgekürzt angegeben werden, vermutlich entsprechend vielverbreiteter wissenschaftlicher Zitierformen (vgl. [Univ. Zürich \(2011a\)](#), 12 und Datensatz [A.5](#)).

<sup>156</sup>[Warner \(2010\)](#), 4.

<sup>157</sup>[Salo \(2009\)](#), 253, weist auf die in OPACs entstehende Diskrepanz zwischen den bibliothekarisch erschlossenen Monographien mit Personennormdaten und den unselbständigen Publikationen ohne solche Normdaten hin.

<sup>158</sup>[Chapman u. a. \(2009\)](#), 324, erläutern knapp die Wichtigkeit einer Kooperation der verschiedenen Repositoriums-Typen.

tensatz einer Normdatei nationalen (in Deutschland die *Personennamendatei (PND)*<sup>159</sup>) oder internationalen (z. B. Virtual International Authority File (*VIAF*)<sup>160</sup>) Charakters kann als eine Verknüpfung genutzt werden, die die Identifizierung über die Grenzen des eigenen Repositoriums hinweg erlaubt. Im Prinzip wäre jeder eindeutig referenzierbare Identifier nutzbar, selbst URLs von *Facebook*- oder *LinkedIn*-Profilen.<sup>161</sup>

Für die Zuordnung des Repositoriums-Namensatzes zu einem übergreifenden Normdatensatz müssen eindeutige Matches ermittelt werden<sup>162</sup>, z. B. anhand der Individualisierungsangaben wie dem Geburtsjahr. Der ermittelte Satz kann dann in dem unter *Anforderung V-1* geschilderten Repositoriumsnamensatz referenziert bzw. bei Datenlieferungen oder in Linked-Data-Anwendungen selbst per URI als Verfasserangabe verwendet werden. In letzterem Fall braucht das Repository keinen eigenen Auflösungsservice seiner Personen-URIs nach außen zu leisten (s. 5.3.2).

Nicht alle Autoren wissenschaftlicher unselbständiger Publikationen sind in Personennamendateien enthalten, da diese, zumindest im bibliothekarischen Bereich, bisher traditionell nur Autoren von Buchveröffentlichungen verzeichnen, denn nur solche werden in Bibliothekskatalogen erschlossen.<sup>163</sup> Der Austausch zwischen Repositorien und Normdateien müsste daher nicht nur zu einer Rücklieferung von URIs für ermittelte Personen führen, sondern auch zur Aufnahme neuer Datensätze für noch nicht vorhandene Personen in die nationale/internationale Normdatei. In den Niederlanden dient zur Referenzierung beispielsweise der *Digitale Auteur Identifier (DAI)*, der von der Wissenschaftsorganisation *SURF* in Zusammenarbeit mit dem *Nationale Thesaurus Auteursnamen* der Hochschulbibliotheken vergeben und von jedem Wissenschaftler beantragt werden kann.<sup>164</sup>

<sup>159</sup>Die kooperativ von den Bibliotheksverbünden und der *DNB* geführte *PND* umfasst normierte Ansetzungen von Namen, die für die bibliothekarische Formal- und Inhaltserschließung benötigt werden. Sie enthält Namensansetzungen sowohl ohne individualisierende Merkmale als auch für individuelle Personen (vgl. Haller u. Fabian (2004), 248ff. Homepage: <<http://www.d-nb.de/standardisierung/normdateien/pnd.htm>>.) Voraussichtlich im Laufe des Jahres 2012 wird die *PND* in die *Gemeinsame Normdatei (GND)* übergehen (vgl. Behrens-Neumann u. Pfeifer (2011)).

<sup>160</sup>Homepage: <<http://viaf.org/>> (Abruf: 10.08.2011).

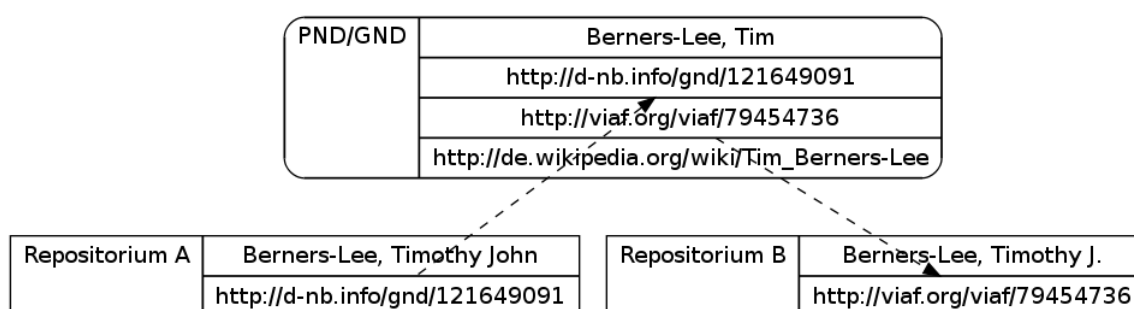
<sup>161</sup>Warner (2010), 3, listet als Beispiel auch kommerzielle Identifier wie die *Scopus Author Id* (<[http://help.scopus.com/robo/projects/schelp/h\\_autsrch\\_intro.htm](http://help.scopus.com/robo/projects/schelp/h_autsrch_intro.htm)> (Abruf: 10.08.2011). Salo (2009), 258f., verweist auf weitere Internetprofildienste für Wissenschaftler, die auch als Identifikation dienen könnten. Auch der Verlagssektor würde von einem einheitlichen, eindeutigen Personentidentifier profitieren und führt zumindest im US-Bereich darüber Gespräche mit bibliothekarischen Anwendern (vgl. CrossRef (2007)).

<sup>162</sup>Solche Verfahren sind direkt während des Submitvorgangs als auch danach denkbar (s. Kap. 6).

<sup>163</sup>Salo (2009), 258; Foulonneau u. André (2008), 60.

<sup>164</sup>Homepage: <<http://www.surffoundation.nl/nl/themas/openonderzoek/infrastructuur/Pages/digitaleauteuridentifizierdai.aspx>> (Abruf: 10.08.2011). Weiteres Beispiel in Brasilien: *Lattes*, eine für einen akademischen Abschluss anstrebende Wissenschaftler obligatorische Perso-

Notwendig sind neben der Gewährleistung einer stabilen und langfristigen Persistenz eines verwendeten Identifiers für Personen auch Verknüpfungen zwischen vorhandenen Systemen. In *VIAF* als internationaler übergreifender Normdatei sind beispielsweise zahlreiche nationale Normdateien referenzierbar integriert. Services wie *OpenID*<sup>165</sup> oder *ISNI*<sup>166</sup> könnten zudem einen übergreifenden Identifier bereitstellen, der auf verschiedene – auch kommerzielle – Einzelsysteme verweist. In Abb. 3 wird als Beispiel gezeigt, wie drei unterschiedlich erfasste Namen der gleichen Person über die Identnummern aus *PND* und *VIAF* miteinander in Verbindung gebracht werden können. Auf diese Weise greifen – alternativ zur für den Submitter aufwändigen Angabe mehrerer Identifier beim Hochladen der Dokumente<sup>167</sup> – Linked-Data-Funktionalitäten. Verschiedene nationale Normdateien können so über ihre internationalen Pendanten Verknüpfungen untereinander und zwischen ihren u. U. abweichenden Namensformen und weiteren Angaben zur Person herstellen.



*Abbildung 3 – Zuordnung verschiedener Namensformen über Normdaten*

Soll die Zuordnung von Personen aus Repositorien zu Normdaten maschinell erfolgen, ist eine standardisierte und erweiterte Erfassung der Namen notwendig: Klare Trennung von Vor- und Nachname, aufgelöste Vornamen, weitere individualisierende Angaben wie Geburtsjahr oder -ort, Fachgebiet etc. Nach der Zuordnung zu einem globalen URI kann

nendatenbank (Homepage: <<http://lattes.cnpq.br/>> [Abruf: 10.08.2011]). Fenner (2011), 29, zählt weitere bereits existierende und sich in Entwicklung befindliche Systeme auf.

<sup>165</sup>Informationen über die *OpenID* sind zu finden unter: <<http://openid.net/get-an-openid/what-is-openid/>> (Abruf: 10.08.2011).

<sup>166</sup>Informationen über den *ISNI* sind zu finden unter <<http://www.isni.org>> (Abruf: 10.08.2011).

<sup>167</sup>Warner (2010), 7.



dieser im Repository genutzt werden, um eventuelle Aktualisierungen und Erweiterungen des Namenssatzes in der Normdatei mit zu berücksichtigen.

### **Anforderung V-3: Die Reihenfolge der Autoren eines Aufsatzes muss in den Metadaten eindeutig erkennbar sein und erhalten bleiben.**

Die Position einer Person in der Autorenliste hat Bedeutung in akademischen Bewerbungsverfahren und Berufungsverfahren, aber auch in kumulativen Dissertationen – oftmals wird eine bestimmte Anzahl von Aufsätzen erwartet, bei denen der Bewerber bzw. Promovend an erster Stelle steht.<sup>168</sup> Diese Position muss in den Metadaten daher bewahrt werden und darf durch keine Prozesse des Metadatenverkehrs, etwa durch XML-Transformationen oder Datenformat-Konversionen, verändert werden.

Viele Datenformate (z. B. DC Simple) legen keine Reihenfolge wiederholter Elemente fest; die Abfolge der Autoren sollte daher in einem dafür geeigneten Element fest abgelegt werden, damit sie immer nachvollziehbar ist.

## **5.4 Sachliche/Inhaltliche Erschließung**

Inhaltliche Erschließung dient dazu, ein Dokument über den Inhalt beschreibende Deskriptoren auffindbar zu machen, da Stichworte aus dem Titel meist zu unspezifisch sind. Dokumente können dabei über Terme aus Thesauri oder über Notationen aus Klassifikationen erschlossen werden.

### **5.4.1 Vorgaben DC Simple/SWAP**

- Kontrolliertes Vokabular oder ein formales Klassifikationssystem verwenden (Empfehlung).<sup>169</sup>
- Schlagwörter/Notationen aus verschiedenen Systemen in getrennten Elementen unterbringen; Schlagwörter/Notationen gleicher Herkunft können in einem gemeinsamen Element mit Semikolon getrennt aufgeführt werden.<sup>170</sup>
- Nennung des verwendeten Klassifikations- bzw. Verschlagwortungssystems<sup>171</sup>

<sup>168</sup>Vgl. z. B. Univ. Gießen (2005), §15, Abs. 3, Satz 1; Univ. Zürich (2010), DO.35.

<sup>169</sup>Hillmann (2005), 4.2.

<sup>170</sup>Hillmann (2005), 4.2.; SWAP (2009).

<sup>171</sup>SWAP (2009).

### 5.4.2 Vorgaben DINI/DRIVER

- Für alle Dokumente eine verbale Sacherschließung durch frei vergebene Schlagwörter oder eine klassifikatorische Erschließung vornehmen.<sup>172</sup>
- Weitere normierte Systeme für die verbale oder klassifikatorische Erschließung (fachübergreifend oder fachspezifisch) verwenden.<sup>173</sup>
- Zusätzlich englische Schlagwörter vergeben.<sup>174</sup>
- Zumindest eine DDC-Sachgruppe entsprechend der Anwendung der DNB (DINI) bzw. aus der vergleichbaren DDC Summary (DRIVER) vergeben.<sup>175</sup>
- Klassifikationen und verbale Schlagwörter getrennt erfassen. Das erste Element der Inhaltserschließung sollte auf jeden Fall ein Schlagwort enthalten.<sup>176</sup>
- Mehrere Schlagwörter mit Semikolon getrennt angeben oder in getrennten Elementen. Klassifikatorische Notationen immer in getrennten Elementen erfassen.<sup>177</sup>
- Die Herkunft des Schlagworts bzw. der Notation muss im Datensatz erkennbar sein, z. B. anhand der Spezifikation für das OAI-PMH-Set. Für Notationen oder kontrollierte Schlagwörter sollten URIs verwendet werden, auch hieraus geht die Herkunft hervor. Notationen sollten im folgenden Element dann verbal aufgelöst werden.<sup>178</sup>

### 5.4.3 Allgemeine Qualitätskriterien

- *accuracy*: Kontrolliertes Vokabular zur Vermeidung von Tippfehlern verwenden.
- *provenance*: Die Herkunft der Inhaltserschließung sollte erkennbar sein: Die Vergabe bspw. von Schlagwörtern durch den Autor des Dokuments ist qualitativ hochwertiger als die Schlagwortvergabe durch ein automatisches Verfahren.

---

<sup>172</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 27 (M.6-3).

<sup>173</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 28 (E.6-1).

<sup>174</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 28 (E.6-2).

<sup>175</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 27 (E.6-4) u. 63 (M.A.3-6); DDC-Sachgruppen der DNB: <[http://www.d-nb.de/service/pdf/ddc\\_wv\\_aktuell.pdf](http://www.d-nb.de/service/pdf/ddc_wv_aktuell.pdf)> (Abruf: 10.08.2011); Vanderfeesten u. a. (2008), 61ff.; DDC Summary: <<http://www.oclc.org/dewey/resources/summaries/>> (Abruf: 10.08.2011).

<sup>176</sup>Vanderfeesten u. a. (2008), 62.

<sup>177</sup>Vanderfeesten u. a. (2008), 62.

<sup>178</sup>Vanderfeesten u. a. (2008), 62 u. 115.

- *conformance to expectations*: Vokabular (eine Systematik oder einen Thesaurus) verwenden, das einen Bezug zur Fachcommunity aufweist.
- *consistency*: Die Verwendung von URIs bei der inhaltlichen Erschließung ermöglichen eine Referenzierung zu Normdateien und Thesauri und damit zu normierten, standardisierten und konsistenten Angaben.
- *currency*: Eine Anpassung ist erforderlich, wenn das Schlagwort- oder Klassifikationssystem sich ändert, ähnlich wie bei einer Umlenkung von Normdaten: Werden beispielsweise zwei vorher getrennte Schlagwörter in einem zusammengefasst, muss dieses in den Daten, die die beiden Schlagwörter verwendet haben, nachvollzogen werden. Am einfachsten geschieht dies durch die Vergabe eines URIs – dieser verweist auch nach der Zusammenlegung immer auf den korrekten Normdatensatz.

#### 5.4.4 Funktionale Anforderungen

##### **Anforderung S-1: Die Objekte sollten über eine dem Fachgebiet gemäße inhaltliche Erschließung gefunden werden können.**

Je nach Fachgebiet sind verschiedene Formen der inhaltlichen Erschließung verbreitet. Spezifische Klassifikationen, die aus gängigen Fachzeitschriften oder -bibliographien stammen, sind dem Submitter vertrauter als bspw. die universelle *Dewey Decimal Classification (DDC)*.<sup>179</sup> Beispiele hierfür sind die *Mathematics Subject Classification (MSC)*<sup>180</sup>, das *Physics and Astronomy Classification Scheme (PACS)*<sup>181</sup> oder das *Journal of Economic Literature (JEL) Classification System*<sup>182</sup>. Die Einordnungsmöglichkeit von Publikationen in solchermaßen bekannte Systematiken erhöht einerseits die Qualität der inhaltlichen Erschließung durch den Submitter, andererseits die Sichtbarkeit der Publikation in fachspezifischen Umgebungen.

##### **Anforderung S-2: Die Objekte sollten über eine konsistente inhaltliche Erschließung gefunden werden können.**

Unterschiedliche Deskriptoren für die selben Zusammenhänge erschweren das Auffinden zusammengehörender Publikationen. Daher sollte die inhaltliche Erschließung durch kon-

<sup>179</sup>Die *DDC* ist eine weitverbreitete bibliothekarische Universalklassifikation, vgl. beispielsweise Taylor u. Joudrey (2009), 78.

<sup>180</sup>Homepage: <<http://www.ams.org/mathscinet/msc/msc.html>> (Abruf: 10.08.2011).

<sup>181</sup>Homepage: <<http://www.aip.org/pacs/>> (Abruf: 10.08.2011).

<sup>182</sup>Homepage: <[http://www.aeaweb.org/journal/jel\\_class\\_system.php](http://www.aeaweb.org/journal/jel_class_system.php)> (Abruf: 10.08.2011).

kontrolliertes Vokabular vorgenommen werden. Hierdurch werden auch einfache Tippfehler vermieden. Zudem macht eine standardisierte Erfassung eine spätere Zuordnung zu anderen Klassifikationssystemen und Thesauri wesentlich einfacher (s. [Anforderung S-4](#)). Freies Vokabular sollte als zusätzliche Beschreibung möglich sein.

**Anforderung S-3: Die Objekte sollten über eine hierarchische inhaltliche Erschließung verfügen, die ein Browsen in unter-, neben- und übergeordneten Klassen ermöglicht.**

Die Einordnung in einen Thesaurus, der über hierarchische Relationen verfügt, bzw. in eine ebensolche Klassifikation, ermöglicht das Browsen nach und die Zusammenfassung von verwandten Publikationen.<sup>183</sup>

**Anforderung S-4: Datensätze sollten unabhängig von der benutzten Sprache und dem genutzten Erschließungssystem inhaltlich gefunden werden können.**

Datensätze aus institutionellen Repositorien sollen global und international genutzt und gefunden werden können, die Verwendung einer bestimmten Sprache oder eines bestimmten, u. U. proprietären Erschließungssystems sollte hierfür keinen Hinderungsgrund darstellen.

Notationen aus Klassifikationen setzen beispielsweise keine Kenntnis einer spezifischen Sprache, dafür aber das Wissen der Bedeutung der Notation voraus. Nicht die zufällige regionale Herkunft des Suchenden, sondern seine fachliche Zugehörigkeit wird so relevant.

Deskriptoren aus Schlagwortsystemen sollten immer auch in einer englischen Übersetzung vorhanden bzw. über ihre englische Form suchbar sein. Auch zu anderen Thesauri und Klassifikationssystemen sollten Verbindungen geknüpft werden können. Hierzu bietet sich der Einsatz von Crosskonkordanzen an. Die Zuordnung von Deskriptoren und Notationen aus verschiedenen Thesauri und Klassifikationen über Crosskonkordanzen dient der Suchraumerweiterung: Dokumente werden unter den Deskriptoren verschiedener Erschließungssysteme gefunden, obwohl sie nur nach einem erschlossen wurden. Auch Alerting-Dienste können Crosskonkordanzen nutzen, um Nachweise aus umfangreichen, heterogen verschlagworteten oder klassifizierten Quellen bieten zu können.<sup>184</sup>

Es existieren bereits Crosskonkordanzen zwischen den verschiedensprachigen *Library of Congress Subject Headings (LCSH)*<sup>185</sup>, *Répertoire d'autorité matière encyclopédique et*

---

<sup>183</sup>Vgl. [Foulonneau u. Riley \(2008\)](#), 98f.

<sup>184</sup>Näheres zu Crosskonkordanzen vgl. [Mayr u. Petras \(2008\)](#).

<sup>185</sup>Die [LCSH](#) umfasst die Liste der von der [LoC](#) verwendeten kontrollierten Schlagworte, die auch von vielen anderen Bibliotheken im anglo-amerikanischen Bereich genutzt wird und somit einen standardisierenden Charakter hat (vgl. [Taylor u. Joudrey \(2009\)](#), 78 u. 349). Die [LCSH](#) werden auch als

*alphabétique unifié* (*RAMEAU*)<sup>186</sup> und *Schlagwortnormdatei* (*SWD*)<sup>187</sup> sowie zwischen der *SWD* und der *DDC*.<sup>188</sup> Um diese Übersetzungsfunktionen nutzen zu können, sollte in den institutionellen Repositorien kontrolliertes Vokabular zur Inhaltserschließung genutzt werden, da eine nachträgliche, maschinelle Zuordnung zu den Termen der Crosskonkordanz ansonsten zu unsicher ist. Nicht nur der Deskriptor/die Notation muss in den Metadaten vorhanden sein, sondern auch die Angabe des zugrundeliegenden Thesaurus/Klassifikationssystems und der Sprache, um in aggregierten Sammlungen mit verschiedenen Erschließungsarten Crosswalks zu ermöglichen. Am empfehlenswertesten ist die Abbildung der Deskriptoren auf URIs, die wiederum auf Datensätze verweisen, die alle notwendigen Informationen enthalten.

## 5.5 Quelle

Die Quellenangabe dient der genauen Identifizierung der Zeitschrift und ihrer Ausgabe, in der das Dokument im Original veröffentlicht wurde. Sie kann als eindeutiger Identifier für das hochgeladene Dokument dienen, wenn es sich dabei um ein offizielles Verlags-PDF handelt, bei Pre- oder Postprints ist sie nur ein Hinweis darauf, an welcher Stelle die zitierfähige Version des Artikels gefunden werden kann.

### 5.5.1 Vorgaben DC Simple/SWAP

- *identifier* oder alternativ *description* zum Transport der Quelle eines Zeitschriftenartikels nutzen.<sup>189</sup> *Mapping SWAP* (2008) schlägt *relation* vor.
- Mindestanforderung an den Inhalt: Zeitschriftentitel oder ein eindeutiger Identifier für die Zeitschrift, Band- und Heftangabe, Anfangs-Seitenzahl/-stelle. Zusätzlich möglichst: End-Seitenangabe; Erscheinungsdatum. Zusätzlich kann auch ein

---

Linked Data zur Verfügung gestellt. Homepage der LCSH: <<http://id.loc.gov/search/>> (Abruf: 10.08.2011).

<sup>186</sup>*RAMEAU* ist eine einheitliche nationale Schlagwortliste für Frankreich (vgl. Mingam (2005). Homepage: <<http://rameau.bnf.fr/>> (Abruf: 10.08.2011).

<sup>187</sup>Projekt Multilingual Access to Subjects (MACS), Homepage: <<https://macs.hoppe.nl/pub/>> (Abruf: 10.08.2011). Die kooperativ von den Verbänden und der DNB geführte *SWD* enthält den terminologisch kontrollierten Wortschatz, der in Deutschland zur Schlagwortvergabe nach den Regeln für den Schlagwortkatalog (RSWK) verwendet wird (vgl. Haller u. Fabian (2004), 252f.). Homepage: <<http://www.d-nb.de/standardisierung/normdateien/swd.htm>> (Abruf: 10.08.2011). Voraussichtlich im Laufe des Jahres 2012 wird die *SWD* in die *GND* übergehen (vgl. Behrens-Neumann u. Pfeifer (2011)).

<sup>188</sup>Projekt CrissCross, Homepage: <<http://www.d-nb.de/wir/projekte/crisscross.htm>> (Abruf: 10.08.2011).

<sup>189</sup>*Apps* (2005).

*OpenURL-ContextObject* (vgl. [Anforderung Q-2](#)) zur Angabe der Quelle angegeben werden.<sup>190</sup>

### 5.5.2 Vorgaben DINI/DRIVER

Weder im *DINI-Zertifikat* noch in den *DRIVER Guidelines* gibt es explizite Hinweise, auf welche Art und Weise die Angabe der Quelle eines Zeitschriftenartikels erfolgen soll. DRIVER verweist in der Beschreibung des DC-Elements *source* auf die Anforderungen der *Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata* ([Apps \(2005\)](#)), die sich primär auf die Beschreibung von „bibliographic citations for journal articles“ fokussieren, beschränkt die Nutzung des Elements allerdings explizit auf Digitalisierungen nicht-digital vorliegender Quellen.<sup>191</sup> Das *DINI-Zertifikat* übernimmt diese Aussage fast identisch.<sup>192</sup> Beide Anforderungspapiere befinden sich damit zwar in Einklang mit dem DC-Simple-Usageguide für *source*<sup>193</sup>, nennen allerdings keine anderen Möglichkeiten für die Angabe einer Quelle, im Gegensatz zu den erwähnten DC Zitier-Guidelines (s. [5.5.1](#)).

In der Praxis nutzen alle untersuchten Repositorien dennoch das DC-Element *source*, um die Quellenangaben zu transportieren (s. z. B. Datensatz [A.2](#)).

### 5.5.3 Allgemeine Qualitätskriterien

- *completeness*: Quellenangabe muss enthalten: Titel der Zeitschrift, Band-/Heftangabe, Jahr, Seitenzahl; muss nicht enthalten: Subskriptionspreis (vielleicht möglich, aber nicht nötig).
- *accuracy*: Eindeutige Zeitschriftentitel bzw. -abkürzungen verwenden; Band- und Heft angeben, wenn anwendbar; kontrolliertes Vokabular (z. B. eine verbindliche Zeitschriftenliste im Hintergrund des Eingabefelds) hilft Tippfehler zu vermeiden.
- *conformance to expectations*: Wenn Abkürzungen für Zeitschriftentitel verwendet werden, die gängigen Konventionen der Community berücksichtigen (z. B. aus entsprechenden Fachbibliographien).

<sup>190</sup>[Apps \(2005\)](#); [SWAP \(2009\)](#) verweist für inhaltliche Anforderungen ebenfalls auf [Apps \(2005\)](#).

<sup>191</sup>[Vanderfeesten u. a. \(2008\)](#), 75.

<sup>192</sup>[DINI-Arbeitsgruppe Elektronisches Publizieren \(2011\)](#), 64: „Das Element dient zur Nennung einer Vorlage für die elektronische Version (Zitatangabe).“

<sup>193</sup>[Hillmann \(2005\)](#), 4.5

- *consistency*: Die Quelle immer nach der gleichen Struktur angeben, möglichst nach einem verbreiteten Standard.
- *consistency*: Die Verwendung von URIs in der Quelle/für den Zeitschriftentitel ermöglichen eine Referenzierung zu Normdateien und damit zu normierten, standardisierten und konsistenten Angaben.
- *timeliness* Eine Anpassung ist erforderlich, wenn ein bisher nur online veröffentlichter Artikel gedruckt und damit unter der Band-/Heftnummer einer Zeitschrift zitierbar wird.

#### 5.5.4 Funktionale Anforderungen

##### **Anforderung Q-1: Die Quellenangabe einer Publikation soll eindeutig einer bestimmten Zeitschrift zuzuordnen sein.**

Für den Zugriff auf die zitierfähige Version einer Publikation, die im Repositorium u. U. als Preprint oder nur unformatiertes Postprint vorhanden ist, ist die genaue Identifikation der Zeitschrift und des Bands/Hefts unabdingbar. Die Zeitschrift kann zudem über eine z. B. durch den Impact Factor definierte Bedeutung verfügen, die für den Submitter in akademischen Bewerbungsverfahren oder Evaluationen wichtig sein kann. Auch für kumulative Dissertationen wird oft eine bestimmte Anzahl an Publikationen in sogenannten „A-Zeitschriften“ oder „B-Zeitschriften“ verlangt.<sup>194</sup>

Zeitschriftentitel sollten daher in einem eigens hierfür definierten Feld und nicht abgekürzt erfasst werden, oder, wenn doch, in einer standardisierten, im speziellen Fachgebiet verbreiteten und bekannten Form. Die Erfassung des Zeitschriftentitels über kontrolliertes Vokabular erhöht die Eindeutigkeit und erleichtert die Zuordnung zu einem Datensatz einer übergeordneten Normdatei von Zeitschriften, der dann per URI referenziert werden kann, oder, falls kein URI existiert, zumindest einen numerischen Identifier (z. B. dieentsprechende Datenbank-ID) liefern kann (s. [Anforderung Q-2](#)).

##### **Anforderung Q-2: Es sollte möglich sein, das Vorhandensein einer Publikation anhand der Quellenangabe in lokalen OPACs oder Bibliotheksportalen anzuzeigen.**

Je nach der Rechercheumgebung (Bibliotheks-OPAC oder -Portal) sollte es möglich sein, mit wenigen Klicks Informationen über Zugriffsmöglichkeiten z. B. des eigenen Biblio-

---

<sup>194</sup>Vgl. z. B. [Univ. Witten \(2005\)](#), [1].



theekssystemen auf die vom Verlag veröffentlichte Version des Artikels (eine gedruckte Ausgabe der Zeitschrift oder den elektronischen Volltext) zu erhalten.

Ein exakt erfasster Zeitschriftentitel (s. [Anforderung Q-1](#)) führt beim Abgleich an den Datensätzen einer übergeordneten Normdatei, in Deutschland beispielsweise der Zeitschriftendatenbank ([ZDB](#))<sup>195</sup>, in der auch Bestandsangaben von Bibliotheken maschinell auswertbar erfasst sind, in den meisten Fällen zu einer eindeutigen Referenz. Diese Referenz kann dann genutzt werden für die gemeinsame Verfügbarkeitsrecherche der [ZDB](#) und der *Elektronischen Zeitschriftenbibliothek* ([EZB](#))<sup>196</sup> auf Grundlage des *OpenURL*-Protokolls, die Ergebnisse auf Heftebene für Bibliotheken zurückliefert.<sup>197</sup> Auch Band- und Heftangaben der Quelle sollten also eindeutig und standardisiert erfasst werden.

## 5.6 Affiliation

Als Affiliation wird im Allgemeinen die Institution bezeichnet, der der Verfasser zum Zeitpunkt der Veröffentlichung des Dokuments angehört (ein Universitäts- oder Forschungsinstitut, ein Lehrstuhl). Relevant ist zusätzlich, dass das veröffentlichte Dokument zum Output der verzeichneten Institution selbst gerechnet werden muss und dies auch weiterhin gilt, wenn der Autor später seine Affiliation wechseln sollte.

### 5.6.1 Vorgaben DC Simple/SWAP

- In [DC Simple](#) gibt es kein explizites Element, in dem die Affiliation untergebracht werden kann. *publisher* verzeichnet die Institution, die für die Veröffentlichung des beschriebenen Dokuments verantwortlich ist, demnach den Betreiber des Repositoriums, das ist aber in den seltensten Fällen der Lehrstuhl oder das Institut des Verfassers, zudem könnten dann nicht die Institutionen der Mit-Autoren erfasst werden. Das Mapping von [SWAP](#), das über ein Element für die Affiliation verfügt, zu [DC Simple](#) nennt hierfür kein mögliches Pendant.<sup>198</sup>

---

<sup>195</sup>Die kooperativ von den Bibliotheksverbänden und der [DNB](#) geführte [ZDB](#) umfasst die überregionalen Nachweise der bibliothekarischen Zeitschriftenbestände in Deutschland (vgl. [Haller u. Fabian \(2004\)](#), 253.)

<sup>196</sup>Die Elektronische Zeitschriftenbibliothek ([EZB](#)) verzeichnet für ihre über 550 teilnehmenden Bibliotheken die Bestandsnachweise und Zugriffsmöglichkeiten wissenschaftlicher elektronischer Zeitschriften (vgl. [Plassmann u. a. \(2006\)](#), 132). Homepage der EZB: <<http://rzblx1.uni-regensburg.de/ezeit/>> (Abruf: 10.08.2011).

<sup>197</sup>Vgl. [Zeitschriftendatenbank \(2011a\)](#).

<sup>198</sup>Mapping [SWAP \(2008\)](#).



- Erfolgt die Angabe des Namens einer Institution mit Hierarchien, sollen die einzelnen Namensbestandteile durch „“ getrennt werden.<sup>199</sup>

### 5.6.2 Vorgaben DINI/DRIVER

Weder im *DINI-Zertifikat* noch in den *DRIVER Guidelines* gibt es Hinweise, auf welche Art und Weise die Angabe der Affiliation des Autors erfolgen soll; DRIVER erwähnt ausschließlich, in welchen DC-Elementen die Affiliation *nicht* erwähnt werden soll: weder in *creator* noch in *publisher*.<sup>200</sup>

### 5.6.3 Allgemeine Qualitätskriterien

- *accuracy*: Kontrolliertes Vokabular verwenden, um Tippfehler zu vermeiden.
- *accuracy*: Universität und/oder Institut eindeutig angeben (nicht „Universität Berlin“, sondern „Humboldt-Universität zu Berlin“).

### 5.6.4 Funktionale Anforderungen

**Anforderung A-1: Die Affiliation des Autors, so wie sie in der Publikation angegeben ist, soll eindeutig erkennbar sein.**

Für Evaluationen, Jahresberichte, Statistiken etc. ist die Angabe der Institution, an der die Publikation erschienen ist, wichtig. Die Affiliation hat als Eigenschaft der Publikation zu gelten, nicht als Eigenschaft des Autors, der seinen Arbeitgeber wechseln kann, wodurch sich auch der unter [Anforderung V-2](#) angesprochene überregionale Normdatensatz ändern würde.<sup>201</sup> Als Affiliation sollte daher die Institution verzeichnet werden, die auch in der offiziellen Verlagspublikation angegeben ist.

IR stellen für die Erfassung der Affiliation oft standardisierte Auswahllisten zur Verfügung, die allerdings nur für die Institution des Submitters Gültigkeit besitzen, nicht für diejenige etwaiger Mitverfasser. Auch freie Angaben sollten allerdings möglich sein, falls

<sup>199</sup>SWAP (2009).

<sup>200</sup>Vanderfeesten u. a. (2008), 61 u. 64: „Usage instruction“ für *publisher*: „The (commercial or non-commercial) publisher of the resource; not the instruction (sub)institution the author is affiliated with.“ In Datensatz [A.2](#) wird dennoch *publisher* für die Angabe der Affiliation genutzt.

<sup>201</sup>Im Idealfall würde der Personennormdatensatz alle Affiliationen chronologisch mit Jahresangaben enthalten, dann wäre eine Zuordnung zur korrekten Affiliation über das Erscheinungsjahr der Publikation möglich – allerdings auch nur bedingt, da sich die Veröffentlichung eines Artikels durch das Peer-Review-Verfahren über Jahre hinziehen kann.

die Publikation aus einer Beschäftigung des Submitters bei einem früheren Arbeitgeber resultierte.<sup>202</sup>

### **Anforderung A-2: Die erfasste Affiliation soll auch über das eigene Repositorium identifizierbar sein.**

Die Erfassung der Affiliation sollte zu einer kontrollierten, allgemeinverständlichen Namensangabe führen. Inneruniversitäre Abkürzungen oder Nummerierungen, die extern keine Bedeutung haben, sollten vermieden werden. Zuordnung zu einer übergeordneten Normdatei und Übernahme eines URIs sind zu empfehlen, da hierüber auch das Finden anderer Namensformen (Verweisungen, Übersetzungen etc.) sowie ein hierarchisches Browsing nach Publikationen von verknüpften unter-/übergeordneten/verwandten Institutionen ermöglicht wird.

## **5.7 Eigenschaften der hochgeladenen Publikation**

Unter diesem Punkt werden Eigenschaften des Dokumentes zusammengefasst, die im Zusammenhang mit seinem Entstehungs- und Publikationsprozess stehen: Versionierung, Begutachtungsstatus, Publikationstyp, Finanzierung etc.

### **5.7.1 Vorgaben DC Simple/SWAP**

Für die Eigenschaften einer unselbständigen Publikation sieht DC Simple keine speziellen Elemente vor, allerdings das auf seiner Grundlage entwickelte SWAP. Hierfür existiert auch ein Mapping zu DC Simple.

- Den Publikationstyp durch einen URI aus dem *Eprints Type Vocabulary Encoding Scheme* angeben.<sup>203</sup>
- Zugriffsberechtigungen durch einen URI aus dem *Eprints AccessRights Vocabulary Encoding Scheme* sowie durch ein Datum, an dem ein eventuell vorhandenes Embargo abläuft, angeben.<sup>204</sup>

---

<sup>202</sup>Grundsätzlich ist zu vermuten, dass ein Wissenschaftler eine gerade erschienene Publikation bei seinem aktuellen Arbeitgeber hochlädt und nicht mehr bei seinem früheren, in dessen Repositorium er u. U. über kein Account mehr verfügt.

<sup>203</sup>Vgl. [http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_Type\\_Vocabulary\\_Encoding\\_Scheme](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Type_Vocabulary_Encoding_Scheme) (Abruf: 10.08.2011) und SWAP (2009).

<sup>204</sup>Vgl. [http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_AccessRights\\_Vocabulary\\_Encoding\\_Scheme](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_AccessRights_Vocabulary_Encoding_Scheme) (Abruf: 10.08.2011) und SWAP (2009).

- Den Begutachtungsstatus durch einen URI aus dem *Eprints Status Vocabulary Encoding Scheme* angeben.<sup>205</sup>
- Geldgebende Forschungsförderungsinstitution/Nummer der Zuwendung angeben.<sup>206</sup>

### 5.7.2 Vorgaben DINI/DRIVER

- *Persistent identifier*<sup>207</sup> bzw. *unique identifier* für jedes Dokument (und jede Version) vergeben.<sup>208</sup>
- Allen Dokumenten Dokument- bzw. Publikationstypen zuweisen, gemäß den Vorgaben aus den DINI-Empfehlungen *Gemeinsames Vokabular für Publikations- und Dokumenttypen* in je eigenen *type*-Elementen, zusätzlich auch einen Typ des *Dublin Core Type Vocabulary*.<sup>209</sup>
- Den Status des Dokuments im Publikationsprozess über OAI-Sets (DINI) bzw. das Element *type* (DRIVER) transportieren.<sup>210</sup>
- Die Reihenfolge der identifier-Elemente innerhalb eines Metadatensatzes so wählen, dass der bevorzugt zu verwendende an erster Stelle steht.<sup>211</sup>
- Die Sprache des Dokuments im Element *language* gemäß der ISO-Norm 639-3 angeben.<sup>212</sup>
- Den Inhalt des Elements *date* gemäß der ISO-Norm 8601 angeben. Das Element soll nur einmal vorkommen und das Publikationsdatum des Dokuments (und nicht das Submitdatum im Repositorium) enthalten.<sup>213</sup>

<sup>205</sup>Vgl. [http://www.ukoln.ac.uk/repositories/digirep/index/Eprints\\_Status\\_Vocabulary\\_Encoding\\_Scheme](http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Status_Vocabulary_Encoding_Scheme) (Abruf: 10.08.2011) und SWAP (2009).

<sup>206</sup>SWAP (2009), kein Mapping auf DC Simple.

<sup>207</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 25 (M.5-10).

<sup>208</sup>Vanderfeesten u. a. (2008), 18 u. 73.

<sup>209</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 63 (M.A.3-5). S. a. 57f. (M.A.2-3, inkl. Liste zu verwendender Benennungen) und 27 (M.6-5); Vanderfeesten u. a. (2008), 68ff. (inkl. Liste kontrollierter Begriffe). *Dublin Core Type Vocabulary*: <http://dublincore.org/documents/dcmi-type-vocabulary/> (Abruf: 10.08.2011).

<sup>210</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 59 (E.A.2-1, inkl. Liste zu verwendender Benennungen); Vanderfeesten u. a. (2008), 68ff. (inkl. Liste kontrollierter Begriffe).

<sup>211</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 64 (E.A.3-1); Vanderfeesten u. a. (2008), 74.

<sup>212</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 63 (M.A.3-7); Vanderfeesten u. a. (2008), 76.

<sup>213</sup>DINI-Arbeitsgruppe Elektronisches Publizieren (2011), 63 (M.A.3-8), 64f. (E.A.3-6); Vanderfeesten u. a. (2008), 66f.

### 5.7.3 Allgemeine Qualitätskriterien

- *consistency*: Erscheinungsjahr: Immer Angabe in der Form „JJJJ“
- *timeliness*: Eine Anpassung ist erforderlich, wenn ein bisher nur online veröffentlichter Artikel gedruckt und damit Band-/Heftnummer einer Zeitschrift zitierbar wird

### 5.7.4 Funktionale Anforderungen

#### **Anforderung P-1: Der Typ der Publikation muss erkennbar sein.**

Für Literaturlisten, Evaluationen und Literatursuchende ist es relevant, ob es sich bei der Publikation um einen wissenschaftlichen Zeitschriftenartikel, eine Research Note, ein Forschungsbericht oder eine Rezension/einen thematischen Literaturüberblick handelt. Für die Angabe sollte auf bereits vorhandene URIs von verbreiteten Ontologien zurückgegriffen werden.

#### **Anforderung P-2: Der Veröffentlichungsstatus des hochgeladenen Dokuments muss erkennbar sein.**

Für die Zitierfähigkeit des hochgeladenen Dokuments soll erkennbar sein, ob es sich um einen Entwurf/Preprint vor Einreichung bei einer Zeitschrift, um die akzeptierte Version (Postprint), um die Verlagsversion etc. handelt. Für die Angabe sollte auf bereits vorhandene URIs von verbreiteten Ontologien zurückgegriffen werden.

#### **Anforderung P-3: Der Begutachtungsstatus der Publikation muss erkennbar sein.**

Für die Bewertung einer Publikation ist es wichtig zu wissen, ob sie ein Peer-Review-Verfahren durchlaufen hat. Für die Angabe sollte auf bereits vorhandene URIs von verbreiteten Ontologien zurückgegriffen werden.

#### **Anforderung P-4: Die Zugänglichkeit des hochgeladenen Dokuments muss erkennbar sein.**

Viele Zeitschriftenverlage erlauben ihren Autoren, Postprints<sup>214</sup> in [IR](#) hochzuladen, diese aber erst ein oder zwei Jahre nach Verlagsveröffentlichung öffentlich zu machen. Während dieses Zeitraums kann das hochgeladene Dokument nur innerhalb der eigenen Institution eingesehen werden. Das Datum, zu dem dieses Embargo abläuft, sollte maschinell auswertbar im Datensatz verankert werden. Nach Auslauf des Embargos kann es theo-

---

<sup>214</sup>Zur Definition eines Postprints s. S. 9 dieser Arbeit.

retisch aus dem Datensatz gelöscht werden, nur die Eigenschaft „öffentlich zugänglich“ ist dann noch relevant.

**Anforderung P-5: Der Zusammenhang der Publikation mit einem geförderten Forschungsprojekt sollte erkennbar sein.**

Viele Publikationen entstehen im Zusammenhang eines Forschungsprojekts, das oft Akteure verschiedener Wissenschaftsinstitutionen umfasst. Um Zusammenhänge zwischen Veröffentlichungen eines Projektes (Deliveries, Reports, Präsentationen, Artikel) herstellen zu können, sollte es in den Metadaten erfasst sein. Eine standardisierte oder referenzierbare Angabe, z. B. über eine eventuell vorhandene ID des Forschungsprojektes in einer Datenbank der betreuenden Forschungsförderinstitution<sup>215</sup> würde es ermöglichen, auch übergreifend vorhandene Dokumente zu finden. Die Angabe einer URL der Projekthomepage wäre ebenfalls denkbar, wobei diese URL wahrscheinlich nicht persistent ist.

**Anforderung P-6: Es muss eindeutig erkennbar sein, mit welcher veröffentlichten Publikation das hochgeladene Dokument in Verbindung steht.**

Suchmaschinen wie *Google Scholar* versuchen, eine Verbindung herzustellen zwischen der veröffentlichten Verlagspublikation und in Repositorien öffentlich zugänglichen Dokumenten. Dem Literatursuchenden kann damit sofort eine – eventuell leicht abweichende – Version einer in seiner Rechercheumgebung nicht lizenzierten Publikation angeboten werden.

Diese Verknüpfung erfolgt am einfachsten über einen eindeutigen und persistenten Identifier der elektronischen Verlagspublikation, der dem veröffentlichenden Wissenschaftler bei Erscheinen der Publikation mitgeteilt wird. Verbreitet sind hier im kommerziellen Bereich vor allem Digital Object Identifiers (DOIs).<sup>216</sup> Daneben existieren u. a. frei erhältliche Uniform Resource Names (URNs)<sup>217</sup> und weitere nicht-kommerzielle

---

<sup>215</sup>Die *Deutsche Forschungsgemeinschaft* betreibt beispielsweise die Datenbank *GEPRIS* für die Verzeichnung geförderter Projekte (Homepage: <<http://gepris.dfg.de/gepris/>> (Abruf: 10.08.2011)).

<sup>216</sup>DOIs werden von allen großen Wissenschaftsverlagen für die elektronischen Versionen ihrer Publikationen und die dazugehörigen Metadaten verwendet sowie oft auch in den Druckerzeugnissen mit angegeben. Die Vergabe von DOIs und die Bereitstellung eines entsprechenden Resolverdienstes erfolgt durch von der DOI-Stiftung autorisierte *registry agencies*; die bekannteste ist *CrossRef*. Zur Entstehung des DOI-Systems und von CrossRef s. [CrossRef \(2009\)](#).

<sup>217</sup>Für den Namensraum *urn:nbn:de* wird beispielsweise von der DNB eine Registrierung und ein Resolverdienst zur Verfügung gestellt (Homepage: <<http://www.nbn-resolving.org/?siteLang=de>> (Abruf: 10.08.2011), mittlerweile werden hier auch DOIs und Handles aufgelöst).

*persistent identifier* aus dem Handle-System, für deren Nutzung eine eher symbolische Jahresgebühr gezahlt werden muss.<sup>218</sup>

#### **Anforderung P-7: Das Objekt soll über einen eindeutigen Identifier referenzierbar sein.**

Das hochgeladene Dokument und die dazugehörigen Metadaten müssen eindeutig identifizierbar sein, um sie anderen Sammlungen oder Umgebungen verwenden zu können. Über diesen Identifier können auch Änderungen an den Metadaten im Original-Repositorium und in aggregierten Sammlungen nachvollzogen werden, die an anderer Stelle vorgenommen wurden, z. B. Anreicherungen durch Normdaten beim Abliefern als Netzpublikation bei der [DNB](#).

### **5.8 Fazit**

Metadaten in [IR](#) werden vor allem bestimmt durch das Datenformat [DC Simple](#), dessen Anwendung ein wichtiger Faktor der Repositorien-Landschaft ist und das dem Austausch der Daten dient. Datensätze sollen aber nicht nur auf strukturell-technischer Ebene austauschbar sein, sondern müssen auch inhaltlich Werte transportieren, die einen Austausch überhaupt sinnvoll machen. Die Inhalte der Metadatenelemente müssen Brücken schlagen zu den Inhalten von Metadaten aus anderen Repositorien oder aus gänzlich anderen Anwendungen in der Wissenschaftskommunikation. Dies ist zu erreichen durch das Einhalten von Standards und die Referenzierung zu Normdaten, so dass Zusammengehörendes leicht erkannt und einander zugeordnet sowie Verwandtes miteinander in Beziehung gesetzt werden kann. Zudem macht dies die Metadaten austauschbar und ihre Verwendung in unterschiedlichen Kontexten möglich, ohne die Notwendigkeit von Kenntnissen, die nur im Ursprungsrepositorium vorhanden sind.<sup>219</sup>

Aus Sicht des Wissenschaftlers ist besonders die eindeutige Identifizierung der Verfasser und der behandelten Thematiken relevant, da nur so Publikationen sicher seinem eigenen Profil hinzugefügt und zielgerichtete Recherchen durchgeführt werden können. Für den unmittelbaren Zugriff auf eine Verlagsversion des Dokuments – durch eine in der lokalen Bibliothek geführte Druckausgabe oder dort lizenzierte elektronische Version – sollte auch die Quelle maschinell auswertbar erfasst werden. Für die Funktionalitäten virtueller Forschungsumgebungen müssen auf jeden Fall auch Publikationseigenschaf-

---

<sup>218</sup>Informationen über *Handles* sind zu finden unter: <<http://www.handle.net>> (Abruf: 10.08.2011) (auch [DOIs](#) gehören zum Handle-System).

<sup>219</sup>Vgl. [Foulonneau u. Riley \(2008\)](#), 176f.

ten wie die verschiedenen Dokumenttypen und -status oder Zugangsberechtigungen klar erkennbar sein.

Standardisierte Eingaben können durch kontrolliertes Vokabular erfolgen, ihr Austausch dann über die entsprechenden Strings (mit Angabe des definierten Vokabulars als *vocabulary encoding scheme*) oder, auf Semantic-Web-Techniken aufsetzend, als URI, falls vom Vokabular zur Verfügung gestellt. Der Einsatz von URIs, der beinahe für jeden zu erfassenden Elementinhalt der Metadatenätze für Zeitschriftenartikel möglich ist, erleichtert auch Rückbezüge zu Veränderungen in der referenzierten Normdatei. Im Kontext von Linked Data schließlich ist dann ein Nachvollziehen von Änderungen nicht mehr notwendig, da die Inhalte der Metadatenelemente erst bei Anzeige aus den referenzierten Datensätzen erzeugt werden, neben den URIs also keine statischen Informationen im Datensatz selbst mehr vorhanden sind. Grundsätzlich sollten die Metadaten nicht nur über [OAI-PMH](#) zur Verfügung gestellt, sondern auch selbst als Linked Open Data, unter Nutzung einer oder mehrerer geeigneten Ontologien, ins Netz gestellt werden.

Die Auswahl der zu benutzenden kontrollierten Vokabulare und/oder Normdateien hängt jeweils von der Ausrichtung des [IR](#) ab. Entscheidend sollte für die Auswahl sein, dass Bezüge zu weiteren Normdateien vorhanden sind, in Form von Crosskonkordanzen oder der Verzeichnung von URIs anderer Herkunft. Metadaten werden auf diese Weise universeller verwendbar und können zukünftig weitere Informationen heranziehen.

## 6 Unterstützungsmöglichkeiten zur Verbesserung der Erschließung selbstarchivierter Dokumente

Die Erschließung selbstarchivierter Dokumente in institutionellen Repositorien erfolgt in der Regel durch den Wissenschaftler bzw. eine von ihm beauftragte Person. Genutzt wird hierzu eine meist repositorienssoftware-spezifische Erfassungsmaske, die die für den jeweilige Publikationstyp relevanten Elemente zum Ausfüllen zur Verfügung stellt. Im IR der Universität Bielefeld, *BiPrints*, handelt es sich beispielsweise bei Zeitschriftenartikeln um 21 Felder.<sup>220</sup> Grundsätzlich wird die Komplexheit der Eingabemaske eines Repositoriums als Hindernis angesehen. McKay (2007) gliedert in ihrer Literaturübersicht über Usability von IR die Probleme u. a. in folgende Bereiche:

- unverständliche Terminologie
- umständlicher Erfassungsvorgang
- zu hohe Ansprüche an die Inhalte der einzelnen Metadatenelemente
- fehlende Funktionalität wie Anzeigen bestimmter Elemente nur für definierte Nutzergruppen
- standardisierte/kontrollierte Eingabe, Verknüpfung mit Normdaten<sup>221</sup>

Wissenschaftler, die hohen intellektuellen Aufwand in die Forschung, das Erstellen eines Artikels und in das dem Peer-Review folgende wiederholte Resubmit-Verfahren investieren, haben kein Interesse an Zeitverlust für administrative Aufgaben.<sup>222</sup> Repositorien sind für sie u. U. wichtige Instrumente, die sie nutzen wollen – wie bei Fachbibliographien oder Bibliothekskatalogen wollen sie aber nicht für ihr Funktionieren verantwortlich sein. Eine Studie, die die Relevanz verschiedener Dienstleistungen für Wissenschaftler analysiert und dabei auch IR mit einschließt, kommt zu dem Schluss:

[...] the key requirement from most researchers' perspectives is for services which are there when they need them, but do not interfere with the creative work at the heart of the research process.<sup>223</sup>

---

<sup>220</sup>Aufruf der Eingabemaske: <<http://repositories.ub.uni-bielefeld.de/biprints/uni/?la=de>> (die Eingabemaske ist mittlerweile [10.08.2011] nicht mehr aufrufbar, da die Software von BiPrints umgestellt wird). S. a. Abb. 1.

<sup>221</sup>Vgl. McKay (2007), Abschn. 'Usability for Data Creators and Maintainers'.

<sup>222</sup>Vgl. Foster u. Gibbons (2005).

<sup>223</sup>Research Information Network (2010), 19.



Für die Akquirierung von Inhalten für IR ist die Kooperation der Wissenschaftler unverzichtbar. Es gilt daher, die Erzeugung von Metadaten auf der einen Seite für den Submitter so leicht und bequem wie möglich zu gestalten, auf der anderen Seite aber die im Kapitel 4 aufgestellten Qualitätsanforderungen erfüllen zu können. Vor allem die Verknüpfung mit kontrolliertem Vokabular durch URIs ist manuell kaum und nur fehlerhaft zu leisten.

Ein IR muss daher verschiedene Hilfsmittel und Instrumente zur Verfügung stellen, um dieses Ziel – das Einbringen einer großen Zahl von Postprints mit möglichst hochwertigen Metadaten – zu erreichen. Hierzu können drei Phasen definiert werden, die im Folgenden näher analysiert werden sollen:

- Instrumente *vor dem Submitvorgang*: dienen meist dazu, die Erschließung durch die Bereitstellung von Fremddaten zu vereinfachen.
- Hilfsmittel und Instrumente *während des Submitvorgangs*: dienen meist dazu, den Erschließungsvorgang zu vereinfachen und standardisierte Eingaben zu erhalten.
- Hilfsmittel und Instrumente/Verfahren *nach dem Submitvorgang*: werden für maschinelle Verknüpfungen zu Normdaten, Thesauri etc. und das Eintragen von URIs genutzt.

## 6.1 vor Submitvorgang

Ein für den Wissenschaftler einfaches Verfahren wäre es, generell Fremddaten für unselbständige Publikationen in das IR einzuspielen (z. B. über *CrossRef*), so dass beim Submitvorgang nur noch der entsprechende Metadatensatz ausgesucht und das Dokument hochgeladen werden muss. Dieses Verfahren ist ähnlich zu dem unter 6.2.2 beschriebenen Importverfahren: In beiden Fällen muss erst eine Suche gestartet werden, um den Datensatz zu finden – entweder im eigenen Repository oder in einer fremden Datenquelle. Wenn langfristige Vereinbarungen mit den Datenlieferanten abgeschlossen wurden, ist eine externe Datenhaltung und dann Abruf *on the fly* vorzuziehen, um nicht die eigene Infrastruktur mit zahlreichen Datensätzen zu belasten, die voraussichtlich nie benötigt werden. Entscheidend ist in beiden Fällen die Qualität der Datensätze, die dann unter der Verwendung weiterer Verfahren angereichert werden müssten.

Für genuin als Open Access erschienene „goldene“ Publikationen gibt es oft Verträge zwischen Verlagen und Repositoriumsbetreibern, um automatisiert Metadaten und

Dokumente hochzuladen.<sup>224</sup> Wissenschaftler müssen sich in diesen Fällen nicht um das Einbringen ins IR kümmern, allerdings handelt es sich hier auch nicht um eine Selbstarchivierung im klassischen „grünen“ Sinne, da die Artikel auf den Verlagsseiten ohnedies frei verfügbar sind.

## 6.2 während Submitvorgang

### 6.2.1 Nutzerprofil

Im IR sollte die Einrichtung von Nutzerprofilen möglich sein, in denen wiederkehrende Angaben hinterlegt werden können. Diese können dann beim Eingeben einer neuen Publikation automatisch erneut verwendet werden bzw. als Vorschlag erscheinen: die Identifikationsnummer oder eine URI aus einer Personennamendatei, individualisierende Angaben zum Verfasser, der Name der Affiliation, die zuletzt verwendeten Schlagwörter oder DDC-Notationen oder auch die Namen der Co-Autoren bisher eingegebener Artikel. Auf diese Weise wird der Nutzeraccount selbst zu einem Normdatensatz und ist den unter seiner Nutzung erschlossenen Publikationen eindeutig zuzuordnen.<sup>225</sup> Ein Nutzerprofil sollte auch über mehrere Identitäten verfügen können, falls ein Lehrstuhlmitarbeiter für verschiedene Wissenschaftler zuständig ist.

### 6.2.2 Import von Datensätzen

In vielen Fällen liegen Metadaten für die hochgeladene Publikation bereits in unterschiedlichen Formen vor. Diese sollten importiert werden können, damit eine bereits vorhandene Erschließung nicht ein weiteres Mal vorgenommen werden muss.

**Import aus eigenem Literaturverwaltungssystem** Für die Liste der eigenen Publikationen, die auf persönlichen Webseiten oder im Lebenslauf Verwendung findet, wird oft ein Literaturverwaltungssystem wie *Citavi* oder *EndNote* bzw. Werkzeuge wie BibTex (in Latex, aber auch mögliches Ausgabeformat von Online-Literaturverwaltungssystemen wie *BibSonomy*, *Mendeley* etc.) genutzt. Diese Datensätze sollten in IR importierbar sein.<sup>226</sup>

**Import von Zitationen** Oft liegen Publikationslisten als reine Text- oder Worddatei vor, in fachspezifischen Zitationsstilen wie *MLA Style*, *Chicago Manual of Style*

---

<sup>224</sup>Dies ist zum Beispiel der Fall bei *Goescholar* (vgl. [Goescholar \(2011\)](#), Abschn. „Externe Daten aus Verlagsvereinbarungen“).

<sup>225</sup>Ein solches System beschreibt [Warner \(2010\)](#), 3, für *arXiv*.

<sup>226</sup>ZORA akzeptiert z. B. BibTex-Format und bietet Unterstützung für Citavi und Endnote (vgl. [Univ. Zürich \(2011c\)](#).)

oder anderen. Diese Referenzen sollten ebenfalls importierbar sein, unter Einbindung von Parsern, die diese „flachen“ Datensätze in strukturierte Formate umwandeln.<sup>227</sup>

**Import aus fremden Systemen** Metadaten für die Publikationen der großen wissenschaftlichen Verlage liegen bereits in *CrossRef* vor, im Bereich Biologie und Medizin zudem in *PubMed*. Über die Angabe einer auf *CrossRef* verweisenden DOI (die dem Verfasser bei Ankündigung des Erscheinens durch den Verlag mitgeteilt wird) oder einer PubMed-ID sollten Datensatzübernahmen möglich sein. DOIs oder PubMed-IDs werden den Autoren üblicherweise beim Erscheinen eines Artikels vom Verlag mitgeteilt.

### 6.2.3 Maschinell aus dem hochgeladenen Dokument erzeugte Metadaten

Beim Hochladen des Dokuments können Metadaten unter Zuhilfenahme von OCR-Prozeduren maschinell erstellt werden. Hierzu kann beispielsweise der typische Aufbau der Titelseite eines Postprints genutzt werden, unter Zuhilfenahme verschiedener Wort- und Namenslisten. Zudem verlangen viele Verlage, bei der Veröffentlichung von Postprints im dazu verwendeten PDF die URL oder DOI der offiziellen Verlagspublikation mit anzugeben; diese kann automatisch aus dem Dokument herausgezogen und dann an den DOI-Auflösungsservice von *CrossRef* zur Übernahme der dazugehörigen Metadaten geschickt werden. Das Hochladen des Dokuments muss in diesem Fall der erste Schritt des Submitvorgangs sein.

Projekte, die in diesem Bereich durchgeführt wurden und deren Ergebnisse genutzt werden könnten:

- RepoMMan, verantwortet von der englischen Forschungsförderungsinstitution Joint Information Systems Committee (JISC)<sup>228</sup>
- IMPACT, ein von der EU finanziertes Projekt zur Vereinfachung von Texterkennung und Massendigitalisierungen.<sup>229</sup> Ergebnisse wurden bereits für das Projekt „Digitalisierung deutscher Dissertationen“ an der Universität Innsbruck genutzt, in dessen Rahmen ca. 210.000 Dissertationen eingescannt und per OCR-Verfahren

<sup>227</sup>Z. B. *FreeCite*, Homepage: <<http://freecite.library.brown.edu/>> (Abruf: 10.08.2011).

<sup>228</sup>Homepage des Projekts: <<http://www.hull.ac.uk/esig/repomman/index.html>> (Abruf: 10.08.2011), vgl. besonders Deliverable D-D13: Green u. Awre (2007).

<sup>229</sup>Homepage des Projekts: <<http://www.impact-project.eu/home/>> (Abruf: 10.08.2011).

Metadaten erzeugt wurden, die an den Daten der DNB abgeglichen werden sollen.<sup>230</sup>

#### 6.2.4 Nutzung von APIs zur Verknüpfung mit Normdaten

Die Verknüpfung mit Normdaten oder Thesauri via IDNs oder URIs ist nur mit maschineller Unterstützung zu realisieren. Hierzu müssen entweder spezifische APIs oder abfragbare Linked-Open-Data-Angebote vorhanden sein. Im Folgenden sind bereits existierende Beispiele aufgeführt

**Verknüpfung mit der SWD über OSWD** Das IR der Universität Freiburg, *FreiDok*, bietet bei der Erstellung der Metadaten eine Suche in der SWD via der vom Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) gehosteten API Online-SWD (OSWD) und die automatische Übernahme der selektierten Deskriptoren an.<sup>231</sup> SWD-Identnummern oder URIs werden dabei allerdings noch nicht mit in den Datensatz übernommen.<sup>232</sup> Da jeder SWD-Term eindeutig ist, kann dieser kontrollierte Deskriptor aber eindeutig referenziert werden. Vorzuziehen wäre allerdings die Übernahme eines URIs in der von der DNB für die SWD zur Verfügung gestellten Form `http://d-nb.info/gnd/<SWD-ID>`.

**Verknüpfung mit VIAF** VIAF bietet eine API und Linked Open Data, über die Personennamen abgefragt werden können und eine Auswahl von Treffern zurückgegeben wird. Diese können dann in einer Auswahlmaske beispielsweise anhand des Geburtsdatum oder der Titel anderer Publikationen des selben Autors ausgewählt und die dazugehörige URI übernommen werden.<sup>233</sup>

---

<sup>230</sup>Homepage des Projekts: <<http://www.uibk.ac.at/ulb/dea/referenzen/dissertationen.html>> (Abruf: 10.08.2011), zudem persönliche Auskunft per Email durch den Leiter des Projektes, Günter Mühlberger.

<sup>231</sup>Das FreiDok-Eingabeformular für Zeitschriftenartikel ist unter der URL <[http://www.freidok.uni-freiburg.de/uni/neu\\_allg.php?la=de&type=2&fak\\_neu=1&anzahl\\_creator\\_name=1&anzahl\\_publisher\\_inst=1&date\\_year=200&description2\\_lang=eng&jahr\\_pruef=200&lic=publ-mit-pod](http://www.freidok.uni-freiburg.de/uni/neu_allg.php?la=de&type=2&fak_neu=1&anzahl_creator_name=1&anzahl_publisher_inst=1&date_year=200&description2_lang=eng&jahr_pruef=200&lic=publ-mit-pod)> (Abruf: 10.08.2011) einsehbar.

<sup>232</sup>Dies kann man z. B. an diesem Datensatz im Format XMetaDissPlus erkennen: <<http://www.freidok.uni-freiburg.de/oai2/oai2.php?verb=GetRecord&metadataPrefix=XMetaDissPlus&identifier=oai:freidok.uni-freiburg.de-opus:1987>> (Abruf: 10.08.2011).

<sup>233</sup>Informationen zur VIAF-API sind zu finden unter: <<http://www.oclc.org/developer/documentation/virtual-international-authority-file-viaf/using-api>> (Abruf: 10.08.2011). Eine Anwendung dieser API lässt sich im Zusammenhang einer Fremddatenübernahme von Amazon-MARC-21-Daten als Webservice betrachten: <<http://amazon.libcat.org/>> (Abruf: 10.08.2011).

## Verknüpfung zu dt. Bibliotheksnormdaten über den Linked Data Service der DNB

Die DNB stellt die Datensätze von PND, SWD und *Gemeinsamer Körperschaftsdatei* (GKD<sup>234</sup> über einen Linked Data Service zur Verfügung.<sup>235</sup> Zudem können sie auch über Search/Retrieve via URL (SRU) abgefragt und in verschiedenen Datenformaten zurückerhalten werden. Über diese Schnittstellen ist es möglich, Verknüpfungen über URIs zu Personennamen, Affiliationen, Schlagwörtern und Affiliations-Orten herzustellen. Zudem enthalten die Datensätze oft noch weitere Angaben wie Identnummern der LoC, Verknüpfungen zu anderen Schlagwortsystemen wie RAMEAU oder LCSH oder klassifikatorische Elemente wie DDC-Notationen.

**Verknüpfung zur ZDB** Auch die ZDB verfügt über eine SRU-Schnittstelle.<sup>236</sup> Hierüber kann die ZDB-ID ermittelt und für die in der Quellenangabe genannten Zeitschrift übernommen werden. Referenzierung zur ZDB dient der Verfügbarkeits-Recherche via OpenURL (s. **Anforderung Q-2**) und stellt Zugriff auf weitere in der ZDB vorhandene Informationen wie ISSN, OCLC-Nummer, Verlagsangaben etc. bereit.

### 6.2.5 Vorschlaglisten in den Eingabefeldern

Bei der Erfassung der Metadaten sollten nach der Eingabe von Wortanfängen dynamische Vorschlaglisten der die bereits im Repository vorhandenen Verfasser(namen), Zeitschriftentitel und Schlagwörter erscheinen. Dies führt zu einer standardisierten Eingabe von Werten. Zudem können so automatisch weitere Felder gefüllt werden, z. B. nach Auswahl eines Zeitschriftentitels die dazugehörigen ISSN und Verlage (oder, falls vorhanden, ZDB-ID). Beispiele:

- **KOPS**: Bei der Eingabe eines Verfassernamens erscheint eine Liste der Konstanzer Universitätsangehörigen.<sup>237</sup>
- **ZORA**: Bei der Eingabe eines Verfassernamens erscheint eine Liste der bereits im Repository vorhandenen Namen, die dazugehörige Affiliation wird allerdings nicht automatisch belegt. Bei der Eingabe eines Zeitschriftentitels erscheint eine

<sup>234</sup>Die kooperativ von den Verbänden und der DNB geführte GKD umfasst Namensansetzungen und Verweisungen für Körperschaften, die für die Formalschließung benötigt werden (vgl. Haller u. Fabian (2004), 251). Homepage: <<http://www.d-nb.de/standardisierung/normdateien/gkd.htm>> (Abruf: 10.08.2011). Voraussichtlich im Laufe des Jahres 2012 wird die GKD in die GND übergehen (vgl. Behrens-Neumann u. Pfeifer (2011)).

<sup>235</sup>S. <[http://www.d-nb.de/hilfe/service/linked\\_data\\_service.htm](http://www.d-nb.de/hilfe/service/linked_data_service.htm)> (Abruf: 10.08.2011).

<sup>236</sup>Vgl. Zeitschriftendatenbank (2011b).

<sup>237</sup>Vgl. KOPS (2011), Abschn. „Autor/in“.

Liste der bereits im Repositorium vorhandenen Titel, bei Auswahl wird ISSN und Verlag ebenfalls übernommen.

### 6.2.6 Validierung der Eingaben

Zumindest für die Syntax der Eingaben sollte eine Validierung vorgenommen werden:

- strukturell falsche Eingaben wie mehrere Namen in einem Verfasserfeld oder Nachnamen ohne Vornamen, nicht-standardisierte Jahresangaben, Wörter in den Seitenangaben etc.
- Prüfung, ob alle Pflichtfelder gefüllt worden sind (Verfasser, Titel, Quelle, ...)
- Korrektheit der Zeichencodierung

Eine semantische Validation ist automatisch nur bei Einsatz von kontrolliertem Vokabular möglich, die oben genannten entsprechenden Verknüpfungsmöglichkeiten per URI erhöhen den Qualitätsstandard dadurch um so mehr.<sup>238</sup>

## 6.3 nach Submitvorgang

Auch nach dem Einbringen der Objekte in das [IR](#) können Metadaten weiterhin verändert und angepasst werden.

### 6.3.1 Manuelle Endredaktion

Manche Repositorien, z. B. [ZORA](#), sehen eine manuelle Endredaktion vor, bevor die Metadaten und das dazugehörige Objekt öffentlich sichtbar werden. Auf diese Weise können grobe Eingabefehler korrigiert und die Urheberrechtssituation geprüft werden. Die Anreicherung der Datensätze durch standardisiertes Vokabular oder das Einfügen von URIs ist manuell allerdings zu aufwändig.

### 6.3.2 Automatisierte Anreicherungen

Anreicherungen können auch an bereits eingebrachten Datensätzen automatisiert vorgenommen werden, um den Submitvorgang selbst nicht zu sehr zu erschweren. Voraussetzung hierfür ist allerdings, dass möglichst standardisierte Eingaben erfasst und der Datensatz zumindest syntaktisch validiert wurde.<sup>239</sup>

---

<sup>238</sup>Zur Validation auch von nutzergenerierten Metadaten vgl. [Braun \(2010\)](#), hier auch die Vorstellung von entsprechenden Software-Tools.

<sup>239</sup>Für eine ausführliche Beschreibung solcher Verfahren vgl. [Hillmann u. a. \(2004\)](#), 2ff.

**Abgleich an Normdateien** Verfasser, inhaltliche Erschließung, die Affiliation, ihr Ort und der Zeitschriftentitel können an Normdateien/kontrolliertem Vokabular abgeglichen und bei eindeutigen Treffern durch die entsprechenden Norm-IDs bzw. URIs ergänzt werden. Ein vergleichbares Verfahren wird zur Zeit bei der [DNB](#) eingerichtet, das bei elektronische Ablieferungen für genannte Personen [PND](#)-Sätze identifizieren und in die Metadaten übernehmen soll.<sup>240</sup> Anders als bei der Nutzung einer entsprechenden API während des Submitvorgangs kann bei einem nachträglichen Abgleich allerdings keine Rückfrage nach einem Geburtsdatum des Verfassers oder dem Verlagsort einer Zeitschrift erfolgen, um Eindeutigkeit herzustellen.

**Abgleich an Crosskonkordanzen** Über den Abgleich an Crosskonkordanzen können weitere Informationen in den Datensatz eingebracht werden, bspw. anderssprachige Schlagwörter<sup>241</sup>, Deskriptoren aus anderen Thesauri oder Notationen verschiedener Klassifikationen.<sup>242</sup>

**Automatische Erzeugung weiterer Deskriptoren** Über das Auslesen der Dokumentinhalte und statistische Verfahren können zusätzliche Deskriptoren erzeugt werden, um das Objekt inhaltlich zu beschreiben. Beispielsweise wurden im Projekt Open-Access-Network (s. S. 19) entsprechende Anreicherungsprozesse zur Illustration entwickelt.<sup>243</sup> Ein weiteres Projekt zur automatischen Erzeugung von [SWD](#)-Deskriptoren war an der Universität Saarbrücken geplant, wurde jedoch von der [DFG](#) abgelehnt.<sup>244</sup> An der [DNB](#) werden ebenfalls automatische Verfahren zur inhaltlichen Erschließung ihrer Netzpublikationen eingerichtet.<sup>245</sup>

## 6.4 Fazit

Die Qualität von Metadaten in [IR](#) kann auf vielfältige Weise erhöht werden, dabei können sowohl automatisierte Verfahren eingesetzt werden wie auch auf ausführlichere/standardisiertere Angaben der Submitter gesetzt werden, was durch die beschriebenen Hilfsmittel erleichtert werden sollte. Vor allem Importmöglichkeiten (vorzugsweise

---

<sup>240</sup>Vgl. [Schöning-Walther \(2011\)](#), 33. Für weitere Beispiele zu nachträglichen Abgleichen und Standardisierungen vgl. [Foulonneau u. Riley \(2008\)](#), 45f. und den Artikel von [Hillmann u. a. \(2004\)](#).

<sup>241</sup>Z. B. via [MACS](#).

<sup>242</sup>Z. B. via Crosskonkordanz [SWD](#) oder den Ergebnissen aus dem CrissCross-Projekt. Zu weiteren möglichen Crosswalks zwischen Thesauri und/oder Klassifikationen vgl. [Mayr u. Petras \(2008\)](#). S. a. Anforderung [Anforderung S-4](#).

<sup>243</sup>Vgl. [Müller u. a. \(2009\)](#), Abschn. 4.1.

<sup>244</sup>Vgl. [Herb \(2009\)](#), über die Ablehnung wird nur im Abstract des Repositoriums berichtet, auf dem die Präsentation liegt.

<sup>245</sup>Vgl. [Schöning-Walther \(2011\)](#).



über die Angabe der DOI) müssen vorhanden sein, um unnötige und zeitaufwändige Schreibarbeit zu vermeiden. Für die inhaltliche Erschließung ist das Spezialwissen des Verfassers jeder automatisierten Erschließung überlegen, daher ist es wichtig, gerade hier den Submitvorgang unkompliziert, aber effektiv zu gestalten. Auch die anderen Angaben sollten so erfolgen, dass wie beschrieben nachträglich maschinelle Verfahren angewendet werden können, um zusätzliche/genauere/kontrollierte Informationen zu erhalten. Dies muss nicht immer zu einem erhöhten intellektuellen Aufwand für den Submitter führen, sondern kann bereits durch eine durchdachte Gestaltung der Eingabemaske erreicht werden:

- Vornamen sollten nie abgekürzt erfasst werden.
- Die Quellenangaben sollten strukturiert nach Titel, Jahr, Band, Heft, Seitenzahl in eigenen Feldern erfasst werden.
- Falls vergeben, sollte immer der DOI erfasst werden, zumindest aber die URL, unter der das Verlags-PDF erhältlich ist.
- Schlagwörter sollten immer in einzelnen Feldern angegeben werden.

Zur Illustration werden in Tabelle 1 die verschiedenen Inhaltsstufen eines Metadatensatzes dargestellt: Zuerst die durch den Import eines *CrossRef*-Datensatzes via DOI erhaltenen Elemente (vgl. auch den von CrossRef gelieferten Originaldatensatz A.6 im Anhang), dann die weiteren Angaben durch den Submitter, die während des Erschließungsvorgangs gewonnen werden, und zuletzt maschinelle Anreicherungen unter Nutzung der in den vorhergehenden Stufen gewonnenen Informationen. Die zweite Stufe ist diejenige, die durch manuelle Eingaben Zeit- und Verständnisaufwand erfordert und daher so einfach und transparent wie möglich gestaltet werden muss. Die Tabelle zeigt, dass es möglich ist, eine Vielzahl von Verknüpfungen zu kontrolliertem Vokabular und Normdaten zu erzeugen, ohne den Ablieferer zu sehr zu belasten: Nur sechs manuelle Angaben sind notwendig, darunter drei rein formale (Sprache, Publikationstyp, Begutachtungsstatus). Die Erfassung der inhaltlichen Erschließung durch den Ersteller der Publikation ist grundsätzlich sinnvoll, da er das beschriebene Gebiet am besten beschreiben kann.<sup>246</sup>

Der Abgleich an Normdaten muss auch zu Neueingaben und Anreicherung von Normdatensätzen führen: Namen/Personen, die noch nicht vorhanden sind, müssen erfasst

---

<sup>246</sup>Allerdings geben Autoren oft bereits bei der Einreichung beim Verlag sog. „Keywords“ an (vgl. z. B. Springer (2011a), Abschn. „Title Page“). Sie werden bei einer Abfrage über die DOI von *CrossRef* allerdings nicht mitgeliefert (s. Datensatz A.6.)



werden (in Tabelle 1 ist dies beim 1. Verfasser der Fall); an vorhandenen Datensätzen sollten Verknüpfungen zu anderen Personen erfolgen, mit denen die Person bereits gemeinsam publiziert hat sowie die Titel oder die DOIs ihrer Publikationen aufgenommen und Affiliationen und behandelte Sachgebiete erfasst werden, möglichst als Verknüpfung zu einer anderen Normdatei. Auf diese Weise werden Personen für die Zukunft leichter abgleichbar. Ein ähnliches Verfahren sollte für Schlagworte entwickelt werden: Deskriptoren für neue Gegebenheiten sollten – unter redaktioneller Bearbeitung – den genutzten Thesauri hinzugefügt werden. Dies hat für beide Seiten Vorteile: Die Repositorien und andere ihre Metadaten nutzende Applikationen können in Zukunft die neuen Deskriptoren und dazugehörige hierarchische Verknüpfungen, die ein Browsing ermöglichen, nutzen, während die Thesauri auf dem aktuellen Stand bleiben. Eine enge Kooperation zwischen Repositorien und Anbietern von standardisierten Normdaten ist hierfür notwendig, die die Entwicklung von Schnittstellen, APIs und Workflows umfasst.

**Tabelle 1** – *Verschiedene Stufen eines Metadatensatzes*

| 1. Stufe: Lieferung von Inhalten durch <i>Crossref</i> (s. Anhang, A.6)   |   |
|---|---|
| <i>Titel</i>  | A Highly Active DNA Polymerase with a Fluorous Core   |
| <i>Verf. 1: Vorname</i><br><i>Verf. 1: Nachname</i>   | Bastian<br>Holzberger   |
| <i>Verf. 2: Vorname</i><br><i>Verf. 2: Nachname</i>   | Marina<br>Rubini  |
| <i>Verf. 3: Vorname</i><br><i>Verf. 3: Nachname</i>   | Heiko M.<br>Möller  |
| <i>Verf. 4: Vorname</i><br><i>Verf. 4: Nachname</i>   | Andreas<br>Marx   |
| <i>Reihenfolge Verf. (1., weitere)</i>  | 1.: Holzberger, weitere: Rubini, Möller, Marx   |
| <i>Zs.: Titel</i><br><i>Zs.: ISSN</i><br><i>Zs.: Erscheinungsdatum d. Hefts</i><br><i>Zs.: Bandnr.</i><br><i>Zs.: Heftnr.</i><br><i>Zs.: Seitenangaben.</i> | Angewandte Chemie International Edition<br>1433-7851<br>2010-02-08<br>49<br>7<br>1324- 1327 |
| <i>Online-Erscheinungsdatum d. Artikels</i><br><i>Dokumenttyp</i>   | 2010-01-13<br>Text  |
| Fortsetzung auf der nächsten Seite  |   |

|   |   |
|---|---|
| <i>Publikationstyp</i>                            | Artikel   |
| <i>URL der Verlags-Version</i>                    | <a href="http://doi.wiley.com/10.1002/anie.200905978">http://doi.wiley.com/10.1002/anie.200905978</a>   |
| <i>DOI der Verlags-Version</i>                    | <a href="http://doi.wiley.com/10.1002/anie.200905978">10.1002/anie.200905978</a>  |
| 2. Stufe: Ergänzung von Inhalten durch Submitter  |   |
| <i>DDC-Sachgruppe</i>                             | 540   |
| <i>Schlagwort frei (engl.)</i>                    | DNA polymerases   |
| <i>Publikation Sprache</i>                        | englisch  |
| <i>Publikationstyp/Version</i>                    | postprint   |
| <i>Begutachtungsstatus</i>                        | peer reviewed   |
| <i>Zugangstatus</i>                               | gesperrt bis 2100-01-01   |
| 3. Stufe: Maschinelle Ergänzung von Inhalten/URIs |   |
| <i>Verf. 1: Geburtsjahr</i>                       | 1984 (a)  |
| <i>Verf. 2: Geburtsjahr</i>                       | 1974 (a)  |
| <i>Verf. 2: PND-URI</i>                           | <a href="http://d-nb.info/gnd/129866083/about/rdf">http://d-nb.info/gnd/129866083/about/rdf</a> (b)   |
| <i>Verf. 2: VIAF-URI</i>                          | <a href="http://viaf.org/viaf/60174000">http://viaf.org/viaf/60174000</a> (b)   |
| <i>Verf. 3: Geburtsjahr</i>                       | 1972 (a)  |
| <i>Verf. 3: PND-URI</i>                           | <a href="http://d-nb.info/gnd/124893759/about/rdf">http://d-nb.info/gnd/124893759/about/rdf</a> (b)   |
| <i>Verf. 3: VIAF-URI</i>                          | <a href="http://viaf.org/viaf/18169122">http://viaf.org/viaf/18169122</a> (b)   |
| <i>Verf. 3: Nationalität</i>                      | DE (b)  |
| <i>Verf. 4: Geburtsjahr</i>                       | 1981 (a)  |
| <i>Verf. 4: PND-URI</i>                           | <a href="http://d-nb.info/gnd/14012926X/about/rdf">http://d-nb.info/gnd/14012926X/about/rdf</a> (b)   |
| <i>Verf. 4: VIAF-URI</i>                          | <a href="http://viaf.org/viaf/103430558/">http://viaf.org/viaf/103430558/</a> (b)   |
| <i>Verf. 4: Beruf/Fachgebiet SWD-URI</i>          | <a href="http://d-nb.info/gnd/4009836-9/about/rdf">http://d-nb.info/gnd/4009836-9/about/rdf</a> (b)   |
| <i>Verf. 4: Beruf/Fachgebiet LCSH-URI</i>         | <a href="http://id.loc.gov/authorities/sh85023033#concept">http://id.loc.gov/authorities/sh85023033#concept</a> (c)                                   |
| <i>Verf. 4: Beruf/Fachgebiet RAMEAU-URI</i>       | <a href="http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11981493p">http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb11981493p</a> (c) |
| <i>Affiliation Publikation</i>                    | Universität Konstanz, Fachbereich Chemie (a)  |
| <i>Affiliation Publikation: URI</i>               | <a href="http://d-nb.info/gnd/5256482-4/about/rdf">http://d-nb.info/gnd/5256482-4/about/rdf</a> (d)   |
| Fortsetzung auf der nächsten Seite                |   |

|                                       |   |
|---------------------------------------|---|
| <i>Affiliation Publikation: URI</i>   | <a href="http://d-nb.info/gnd/2003070-8/about/rdf">http://d-nb.info/gnd/2003070-8/about/rdf</a> (d)   |
| <i>DDC-Sachgruppe: URI</i>            | <a href="http://info.eu-repo/classification/ddc/540">info:eu-repo/classification/ddc/540</a> (e)  |
| <i>Schlagwort: LCSH-URI</i>           | <a href="http://id.loc.gov/authorities/sh85037007#concept">http://id.loc.gov/authorities/sh85037007#concept</a> (c)                                   |
| <i>Schlagwort: SWD-URI</i>            | <a href="http://d-nb.info/gnd/4381045-7/about/rdf">http://d-nb.info/gnd/4381045-7/about/rdf</a> (c)   |
| <i>Schlagwort: SWD-URI</i>            | <a href="http://d-nb.info/gnd/4150345-4/about/rdf">http://d-nb.info/gnd/4150345-4/about/rdf</a> (c)   |
| <i>Schlagwort: RAMEAU-URI</i>         | <a href="http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb12265191n">http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb12265191n</a> (c) |
| <i>Schlagwort: DDC-Notation</i>       | <a href="#">572.786</a> (c)   |
| <i>ZDB-ID</i>                         | <a href="#">123227-7</a> (f)  |
| <i>Zs. Verlag</i>                     | <a href="#">Wiley-VCH</a> (f)   |
| <i>Zs. Verlagsort</i>                 | <a href="#">Weinheim</a> (f)  |
| <i>Publikationstyp: URI</i>           | <a href="http://info.eu-repo/semantics/article">info:eu-repo/semantics/article</a> (e)  |
| <i>Begutachtungsstatus: URI</i>       | <a href="http://info.eu-repo/semantics/acceptedVersion">info:eu-repo/semantics/acceptedVersion</a> (e)  |
| <i>URN der IR-Version</i>             | <a href="http://nbn-resolving.org/urn:nbn:de:bsz:352-opus-102908">urn:nbn:de:bsz:352-opus-102908</a> (g)  |
| <i>Emailadresse für Kopie-Anfrage</i> | <a href="mailto:xyz@xyz.de">xyz@xyz.de</a> (a)  |

- 
- (a) Via Profil des Submitters bzw. der Mitautoren der gleichen Institution  
(b) Via [PND](#) / Linked-Data-Service der [DNB](#) (Beispiel s. Anhang A, [A.7](#)).  
(c) Via [LCSH](#) / [SWD](#) / Linked-Data-Service der [DNB](#) / [MACS](#) (Beispiele s. Anhang A, [A.9](#) und Anhang A, [A.8](#)).  
(d) Via [GKD](#) / Linked-Data-Service der [DNB](#) (Beispiel s. Anhang A, [A.10](#)).  
(e) Via hinterlegte Listen im [IR](#), basierend auf [Vanderfeesten u. a. \(2008\)](#), 115ff.  
(f) Via [ZDB](#).  
(g) Via [IR](#) / vom URN-Service der [DNB](#) vergebener Namespace

## 7 Schlussbetrachtung

Die Einrichtung von institutionellen und disziplinären Repositorien gehört neben der Gründung von freien Online-Zeitschriften zu einer der beiden Strategien der Open-Access-Bewegung, um wissenschaftliche Publikationen und Forschungsergebnisse weltweit frei zugänglich zu machen. Unterstützt durch Forschungsförderungsinstitutionen wurden in den letzten Jahren weltweit über 1.500 [IR](#) und mehr als 225 disziplinäre Repositorien eingerichtet – von universitären Rechenzentren, Bibliotheken oder anderen in der Forschung tätigen Einrichtungen.

Für viele Wissenschaftler aber scheint Open Access nicht genügend Relevanz zu besitzen, um den üblichen Schritten des Publikationsprozesses – Forschen, Schreiben, Einreichen, Peer-Review-Verfahren – einen weiteren, nämlich Bereitstellung der erschienenen Publikation bzw. eines Postprints in einem [IR](#), hinzuzufügen. Repositorien sind von Wissenschaftlern bisher nicht als ein wichtiges Instrument anerkannt worden, weder im allgemeinen Wissenschaftsprozess noch im eigenen Arbeitsalltag. Repositorien generieren für sie bisher, im Vergleich zur Reputation versprechenden Veröffentlichung eines Artikel in einer Peer-Review-Zeitschrift, kaum einen Mehrwert.

Diese Arbeit beschäftigt sich damit, Wege zu finden, durch die Dienstleistungen von [IR](#) für die Wissenschaft verbessert oder überhaupt erst generiert werden können. Entscheidend dabei ist, Anforderungen unter Wissenschaftlern darüber zu erheben, welche grundsätzlichen Services sie in ihrer Arbeit unterstützen könnten und welcher Mehrwert für sie überhaupt relevant wäre. Die ausgewerteten Umfragen ergeben in diesem Zusammenhang folgendes Bild:

- Wissenschaftler forschen und kooperieren nicht institutionsbezogen, sondern fachbezogen und global. In Bezug auf Veröffentlichungen fühlen sie sich eher an ihre Fachcommunity gebunden als an ihre Heimatinstitution. Auch Publikationen aus [IR](#) müssen daher in fachlichen Zusammenhängen präsentiert werden können. Weniger die Institutionsbezogenheit der [IR](#), sondern seine technischen Möglichkeiten und Vorteile für die Verbreitung von Literatur sollten daher im Vordergrund stehen.
- Die Reputation und damit auch die Karrierechancen von Wissenschaftlern hängen von der Bedeutung der Zeitschriften ab, in denen sie publizieren. Die Quelle einer Publikation und die Tatsache, dass sie einer Begutachtung unterlag, muss aus der Präsentation im [IR](#), aber auch in anderen Zusammenhängen hervorgehen können.

- Sichtbarkeit ist ein wichtiges Kriterium: Publikationen aus [IR](#) müssen in Suchmaschinen wie *Google* oder *Google Scholar*, Publikationslisten, Alerting-Diensten, RSS-Feeds, Online-Bibliographien etc. einbindbar sein.

Für die Nutzung in virtuellen Forschungsumgebungen ist der institutionelle Charakter der Repositorien ebenfalls irrelevant, hier können sie als international verteilte Zugangspunkte zu Publikationen dienen, aber auch als eine Arbeitsplattform für kollaborativ arbeitende Wissenschaftler verschiedener Institutionen, die Versionsverwaltungen und Unterstützungswerkzeuge für das Einreich- und Begutachtungsverfahren zur Verfügung stellt. Ein weiteres großes Entwicklungspotenzial besteht im Bereich der Bereitstellung von Forschungsrohdaten, die in geeigneter Weise mit den sie auswertenden Publikationen verknüpft und für die Strategien der Langzeitarchivierung entwickelt werden müssen.

Der Sinn und Zweck von [IR](#) besteht daher primär in der Rolle, die sie für den Wissenschaftsprozess spielen. Weitere Motive für ihre Einrichtung wie die Erzeugung von akademischen Jahresberichten oder der Nutzung als „Schaufenster“ der universitären Leistungen stellen für Wissenschaftler hingegen kaum Anreize dar, ihre Publikationen hochzuladen. Im Mittelpunkt stehen die Funktionen der [IR](#), nicht die einzelnen Implementationen. Erfüllen können sie diese nur als Netzwerk, eng verknüpft über standardisierte Schnittstellen, über die sie ihre Daten anderen Sammlungen zur Verfügung stellen und auch Fremdanwendungen Zugriff erlauben. Bereits die [BOAI](#) verkündete: „When these archives conform to standards created by the Open Archives Initiative, then search engines and other tools can treat the separate archives as one.“<sup>247</sup>

Um dieses Ziel zu erreichen, ist aber nicht nur eine standardisierte technische Infrastruktur notwendig, sondern müssen auch die Metadaten, die die Dokumente beschreiben, entsprechende Ansprüche erfüllen, denn sie stellen die Verknüpfungsmöglichkeiten erst her. Neben der Analyse allgemeiner Qualitätskriterien wie Korrektheit oder Vollständigkeit und deren Anwendung auf die einzelnen Metadatenelemente, wurden in dieser Arbeit am Beispiel der Elemente für Zeitschriftenartikel vor allem funktionale Anforderungen herausgearbeitet, die erfüllt sein müssen, damit Repositorien neue Services erbringen sowie integraler Teil der Wissenschaftskommunikation werden können:

- Verfasser müssen eindeutig identifiziert werden können.

---

<sup>247</sup>[BOAI](#) (2002), Abschn. ‘I. Self-Archiving’ (Herv. d. Verf.).

- Die inhaltliche Erschließung muss Crosswalks zulassen, d.h. Verbindungen zu anderen Thesauri oder Klassifikationssystemen, um auch heterogen oder verschieden-sprachig beschriebene Publikationen gemeinsam präsentieren zu können.
- Es muss eine Verbindung zu der Zeitschrift herstellbar sein, in der der Artikel erschienen ist.
- Publikationsstatus und -typen – Artikel, Begutachtung, Pre-/Postprint/Verlags-PDF – müssen eindeutig beschrieben werden.

Die Verwendung von kontrolliertem Vokabular, eindeutigen Identifiern für Personen, Standardnummern für Zeitschriften etc. ermöglicht eine eindeutige Zuordnung zu übergeordneten Normdateien, Thesauri, Klassifikationssystemen und Crosskonkordanzen. Aus diesen können direkt standardisierte Ansetzungen übernommen werden oder referenzierbare URIs, über die Informationen *on the fly* herangezogen werden können. Letztere haben den Vorteil, dass der Metadatensatz dynamisch bleibt, ohne dass er selbst geändert werden muss: Ändert sich eine Ansetzung im Referenzsystem, bleibt die übernommene URI dennoch gültig, obwohl sie von nun an auf den neuen Inhalt verweist.

Anspruchsvollere funktionale Anforderungen an Metadaten bedeuten auch höheren Aufwand für die Erschließung der Dokumente. In [IR](#) wird diese in den meisten Fällen durch die Ablieferer geleistet, d.h. durch Wissenschaftler. Im letzten Kapitel wurden Hilfsmittel und Verfahren beschrieben, die den Vorgang der Metadatenerstellung unterstützen können. Bereits vorliegende Erschließungen jedweder Art – in *CrossRef*, aus eigenen Literaturverwaltungssystemen, aus einfachen Literaturlisten – sollten nutzbar sein, um Doppelarbeit zu vermeiden. Dynamische Vorschlagslisten für im [IR](#) bereits vorhandene Personen oder leicht zu bedienende APIs für die Eingabe von kontrolliertem Vokabular der inhaltlichen Erschließung erzeugen standardisierte Eingaben und erleichtern die Erfassung. Grundsätzlich muss dabei das Ziel sein, die Ablieferer soweit wie möglich vom Verwaltungsaufwand der Metadatenerzeugung zu befreien. Es wurde illustriert, dass die Kombination von Fremddatenübernahmen aus *CrossRef* (via [DOI](#)) und darauf folgende maschinelle Verfahren nur noch wenige zusätzliche manuelle Angaben erfordern, um Abgleiche an nationalen und internationalen Normdateien durchführen zu können. Diese stellen in vielen Fällen APIs oder auch Linked Open Data für Abfragen zur Verfügung und liefern als Ergebnis autorisierte Ansetzungen und URIs zurück. Auch Normdateien sollten sich aber öffnen für Übernahmen aus Repositorien, z. B. von noch nicht vorhandenen Namenssätzen oder Deskriptoren. Hierfür werden redaktionelle und automatisierte Workflows entwickelt werden müssen.

Um die Attraktivität von Repositorien zukünftig zu erhöhen, sind weitere – quantitative und qualitative – Erhebungen unter Wissenschaftlern darüber notwendig, welche Services sie als hilfreich empfänden. Die in dieser Arbeit genutzten Umfrageergebnisse der Projekte *PEER* und *SOAP* (s. S. 11) geben bereits einige Hinweise, sind jedoch vergleichsweise allgemein gehalten, so dass spezielle Befragungen von Wissenschaftlern unter Berücksichtigung verschiedener Fachdisziplinen oder auch Herkunftsländer besser belegbare Ergebnisse erbringen könnten.

Weiterhin wären Wege zu untersuchen, auf welche Art und Weise durch URIs und kontrolliertes Vokabular angereicherte Datensätze am effektivsten auszutauschen wäre: Der Mindeststandard *DC Simple* von *OAI-PMH* ist hierfür kaum geeignet, da er keine Qualifier zulässt, die Aussagen über die Art der jeweiligen Elementinhalte machen können. Um Metadaten für die Erzeugung weiterer Services nutzen zu können, sind daher besser geeignete Datenformate notwendig, wie z. B. *Qualified DC* oder *MODS*<sup>248</sup>. Die bisher gültige Gleichung, dass nur ein einfaches Datenformat auch eine nutzerfreundliche Datenerfassung ermögliche<sup>249</sup>, sollte dabei durch einen umfassenden Einsatz maschineller Hilfsmittel, wie in Kap. 6 beschrieben, so weit wie möglich relativiert werden.

Zusätzlich zum Datenaustausch über *OAI-PMH* wären aber auch andere Bereitstellungsszenarien für die Metadaten denkbar, beispielsweise als *Linked Open Data*. Ihre Repräsentation in *RDF* unter der Verwendung von Ontologien würde die Inhalte von *IR* zu einem Teil des *Semantic Web* machen und weitere Nutzungsmöglichkeiten eröffnen.

Die Vernetzung, die Repositorien leisten sollen, ist ein wichtiger Aspekt der Metadaten selbst. Auch sie dürfen nicht isoliert und statisch bleiben, sondern müssen, bis hinunter auf die Elementebene, verknüpfbar sein und verknüpfend wirken sowie inhaltlich Beziehungen schaffen zu anderen Konzepten und Modellen. Auf diese Weise sind sie flexibel einsetzbar, in Suchmaschinen und Web-/Bibliotheksportalen, im *Semantic Web*, in aggregierten Sammlungen. Repositorien erfüllen dann beide Anforderungen von *Open Access* – *Offenheit* durch die freie Bereitstellungen wissenschaftlicher Publikationen sowie *Zugang* durch eine in vielfältigen Umgebungen nutz- und einbindbare Beschreibung ihrer Inhalte.

---

<sup>248</sup>Vgl. Foulonneau u. André (2008), 28ff, hier auch eine Liste und Kurzbeschreibungen gängiger Datenformate.

<sup>249</sup>So z. B. Foulonneau u. André (2008), 29. „The user-friendliness and simplicity of deposit interfaces are a key success factor for repositories. As a consequence, the metadata formats adopted have to be simple (e.g. similar to DC).“

# Literatur

## Allianz der deutschen Wissenschaftsorganisationen 2011a

ALLIANZ DER DEUTSCHEN WISSENSCHAFTSORGANISATIONEN: *Definition Virtuelle Forschungsumgebung*. <[http://www.allianzinitiative.de/de/handlungsfelder/virtuelle\\_forschungsumgebungen/definition/](http://www.allianzinitiative.de/de/handlungsfelder/virtuelle_forschungsumgebungen/definition/)>.

Version: Februar 2011, Abruf: 10.08.2011

## Allianz der deutschen Wissenschaftsorganisationen 2011b

ALLIANZ DER DEUTSCHEN WISSENSCHAFTSORGANISATIONEN: *Frequently asked Questions zu Open Access und Zweitveröffentlichungsrecht (FAQ)*. 2011 <[http://www.allianzinitiative.de/fileadmin/user\\_upload/FAQ\\_Open\\_Access\\_Zweitveroeffentlichungsrecht.pdf](http://www.allianzinitiative.de/fileadmin/user_upload/FAQ_Open_Access_Zweitveroeffentlichungsrecht.pdf)>, Abruf: 10.08.2011

## Allianz der deutschen Wissenschaftsorganisationen 2011c

ALLIANZ DER DEUTSCHEN WISSENSCHAFTSORGANISATIONEN: *Schwerpunktinitiative „Digitale Information“, Handlungsfelder*. <<http://www.allianzinitiative.de/de/handlungsfelder/>>. Version: 2011, Abruf: 10.08.2011

## Apps 2005

APPS, Anne: *Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata*. <<http://dublincore.org/documents/2005/06/13/dc-citation-guidelines/>>. Version: Juni 2005, Abruf: 10.08.2011 (DCMI Recommendation)

## Behrens-Neumann u. Pfeifer 2011

BEHRENS-NEUMANN, Renate ; PFEIFER, Barbara: Die Gemeinsame Normdatei - ein Kooperationsprojekt. In: *Dialog mit Bibliotheken* (2011), Nr. 1, 37–40. <[http://www.d-nb.de/standardisierung/pdf/dialog2011\\_1\\_37-40\\_behrens\\_pfeifer.pdf](http://www.d-nb.de/standardisierung/pdf/dialog2011_1_37-40_behrens_pfeifer.pdf)>, Abruf: 10.08.2011

## Berlin Declaration on Open Access 2003

*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. <<http://oa.mpg.de/lang/en/berlin-prozess/berliner-erklarung/>>. Version: 2003, Abruf: 10.08.2011. – (Conference on Open Access to Knowledge in the Sciences and Humanities, October 20 – 22, 2003, Berlin)

## Björk u. a. 2010

BJÖRK, Bo-Christer ; WELLING, Patrik ; LAAKSO, Mikael ; MAJLENDER, Peter



; HEDLUND, Turid ; GUÐNASON, Guðni: Open Access to the Scientific Journal Literature: Situation 2009. In: *PLoS ONE* 5 (2010), Nr. 6, e11273. <<http://dx.doi.org/10.1371/journal.pone.0011273>>, Abruf: 10.08.2011

#### **BOAI 2002**

*Budapest Open Access Initiative*. <<http://www.soros.org/openaccess/read.shtml>>. Version: 2002, Abruf: 10.08.2011

#### **Boosen 2008**

BOOSEN, Martin: *Institutionelle Repositorien in Deutschland*, Fachhochschule Köln (Institut für Informationswissenschaft), Master's Thesis, 2008. <<http://hdl.handle.net/10760/12894>>, Abruf: 10.08.2011

#### **Braun 2010**

BRAUN, Jürgen: *Metadatenqualität und -interoperabilität*. <[www.kim-forum.org/material/pdf/KIM\\_BID2010\\_Braun.pdf](http://www.kim-forum.org/material/pdf/KIM_BID2010_Braun.pdf)>. Version: 2010, Abruf: 10.08.2011. – (4. Leipziger Kongress für Information und Bibliothek, 16. März 2010)

#### **Bruce u. Hillmann 2004**

BRUCE, Thomas R. ; HILLMANN, Diane I.: The Continuum of Metadata Quality: Defining, Expressing, Exploiting. In: HILLMAN, Diane I. (Hrsg.) ; WESTBROOKS, Elaine L. (Hrsg.): *Metadata in Practice*. Chicago : American Library Association, 2004, 238–256. – (Postprint unter: <[http://ecommons.cornell.edu/bitstream/1813/7895/1/Bruce\\_Hillmann\\_corr\\_final.doc](http://ecommons.cornell.edu/bitstream/1813/7895/1/Bruce_Hillmann_corr_final.doc)> [Abruf:10.08.2011])

#### **Bruch u. Lengenfelder 2009**

BRUCH, Christoph ; LENGENFELDER, Anja: Unterstützung des Grünen Weges zu Open Access an der Max-Planck-Gesellschaft. In: *GMS Medizin - Bibliothek - Information* 9 (2009), Nr. 1, Doc06. <<http://www.egms.de/en/journals/mbi/2009-9/mbi000134.shtml>>, Abruf: 10.08.2011

#### **Chapman u. a. 2009**

CHAPMAN, John ; REYNOLDS, David ; SHREEVES, Sarah: Repository Metadata: Approaches and Challenges. In: *Cataloging & Classification Quarterly* 47 (2009), Nr. 3, 309–325. <<http://dx.doi.org/10.1080/01639370902735020>>, Abruf: 10.08.2011

#### **Connor 2011**

CONNOR, James: *Google Scholar Citations*. <<http://googlescholar.blogspot>.

[com/2011/07/google-scholar-citations.html](http://com/2011/07/google-scholar-citations.html)>. Version: 2011 (Google Scholar Blog ; 20. Juli 2011)

### **CrossRef 2007**

CROSSREF: *CrossTech: CrossRef Author ID meeting*. <[http://www.crossref.org/CrossTech/2007/02/crossref\\_author\\_id\\_meeting.html](http://www.crossref.org/CrossTech/2007/02/crossref_author_id_meeting.html)>. Version: 2007, Abruf: 10.08.2011

### **CrossRef 2009**

CROSSREF (Hrsg.): *The Formation of CrossRef: A Short History*. Lynnfield/Oxford : CrossRef, 2009 <<http://www.crossref.org/08downloads/CrossRef10Years.pdf>>, Abruf: 10.08.2011

### **Dallmeier-Tiessen u. a. 2011a**

DALLMEIER-TIESSEN, Suenje ; DARBY, Robert ; GOERNER, Bettina ; HYPPOELAE, Jenni ; IGO-KEMENES, Peter ; KAHN, Deborah ; LAMBERT, Simon ; LENGFELDER, Anja ; LEONARD, Chris ; MELE, Salvatore ; NOWICKA, Malgorzata ; POLYDORATOU, Panayiota ; ROSS, David ; RUIZ-PEREZ, Sergio ; SCHIMMER, Ralf ; SWAISLAND, Mark ; STELT, Wim van d.: *Highlights from the SOAP Project Survey: What Scientists Think about Open Access Publishing*. <<http://arxiv.org/abs/1101.5260>>. Version: 2011, Abruf: 10.08.2011

### **Dallmeier-Tiessen u. a. 2011b**

DALLMEIER-TIESSEN, Suenje ; DARBY, Robert ; GOERNER, Bettina ; HYPPOELAE, Jenni ; IGO-KEMENES, Peter ; KAHN, Deborah ; LAMBERT, Simon ; LENGFELDER, Anja ; LEONARD, Chris ; MELE, Salvatore ; NOWICKA, Malgorzata ; POLYDORATOU, Panayiota ; ROSS, David ; RUIZ-PEREZ, Sergio ; SCHIMMER, Ralf ; SWAISLAND, Mark ; STELT, Wim van d.: *What Scientists Think about Open Access Publishing*. <<http://www.slideshare.net/ProjectSoap/soap-symposiumtalkii>>. Version: 2011, Abruf: 10.08.2011. – (The SOAP Symposium, Berlin, 13 January 2011)

### **Davis u. Connolly 2007**

DAVIS, Philip M. ; CONNOLLY, Matthew J. L.: Institutional Repositories: Evaluating the Reasons for Non-use of Cornell University's Installation of DSpace. In: *D-Lib Magazine* 13 (2007), Nr. 3/4. <<http://dx.doi.org/10.1045/march2007-davis>>, Abruf: 10.08.2011

### **Deutsche Forschungsgemeinschaft 2005**

DEUTSCHE FORSCHUNGSGEMEINSCHAFT: *DFG-Positionspapier: Elektronisches Publizieren; Empfehlungen des Unterausschusses für elektronische Publikationen und des Ausschusses für wissenschaftliche Bibliotheken und Informationssysteme*. 2005 <[http://www.dfg.de/download/pdf/foerderung/programme/lis/pos\\_papier\\_elektron\\_publizieren\\_0504.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/pos_papier_elektron_publizieren_0504.pdf)>, Abruf: 10.08.2011

### **Deutsche Forschungsgemeinschaft 2009**

DEUTSCHE FORSCHUNGSGEMEINSCHAFT: *Aufforderung zur Antragstellung: DFG-Förderprogramm Elektronische Publikationen, Aktionslinie „Repositorien – Ausbau und Entwicklung vernetzter Informationsdienstleistungen für die Wissenschaft“*. 2009 <[http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung\\_repositorien\\_100222.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung_repositorien_100222.pdf)>, Abruf: 10.08.2011

### **Deutsche Forschungsgemeinschaft 2010**

DEUTSCHE FORSCHUNGSGEMEINSCHAFT: *Merkblatt Open Access Publizieren*. 2010 <[http://www.dfg.de/download/formulare/12\\_20/12\\_20.pdf](http://www.dfg.de/download/formulare/12_20/12_20.pdf)>, Abruf: 09.08.2011

### **Deutsche Nationalbibliothek 2010**

DEUTSCHE NATIONALBIBLIOTHEK: *XMetaDissPlus: Format des Metadatensatzes der Deutschen Nationalbibliothek für Online-Hochschulschriften inklusive Angaben zum Autor (XMetaPers)*. 2010 <<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:101-2010052704>>, Abruf: 10.08.2011

### **DINI 2011**

DINI: *DINI-Historie*. <<http://www.dini.de/ueber-dini/dini-historie/>>. Version: 2011, Abruf: 09.08.2011

### **DINI-Arbeitsgruppe Elektronisches Publizieren 2011**

DINI-ARBEITSGRUPPE ELEKTRONISCHES PUBLIZIEREN: *DINI-Schriften*. Bd. 3: *DINI-Zertifikat Dokumenten- und Publikationsservice 2010*. Vers. 3.1, März 2011. 2011 <<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:kobv:11-100182794>>, Abruf: 10.08.2011

### **DRIVER 2008**

DRIVER: *About DRIVER*. <<http://www.driver-repository.eu/Driver-About/About-DRIVER.html>>. Version: September 2008, Abruf: 10.08.2011

### European Commission 2011

EUROPEAN COMMISSION: *Research Infrastructures : ESFRI*. <[http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri)>. Version: 2011, Abruf: 10.08.2011

### Eversberg 2003

EVERSBERG, Bernhard: Grundsätze und Ziele des Katalogisierens: Zur Theorie der Bibliothekskataloge und Suchmaschinen. In: BENKERT, Hannelore (Hrsg.): *Die Bibliothek zwischen Autor und Leser: 92. Deutscher Bibliothekartag in Augsburg 2002*. Frankfurt am Main : Klostermann, 2003, 113–126. – (Text ist in leicht abgewandelter Form unter <<http://www.biblio.tu-bs.de/allegro-test/formate/tks.htm#tab>> online erhältlich [Abruf: 10.08.2011])

### Fenner 2011

FENNER, Martin: Author Identifier Overview. In: *LIBREAS* (2011), Nr. 18. <<http://edoc.hu-berlin.de/docviews/abstract.php?id=37867>>, Abruf: 10.08.2011

### Foster u. Gibbons 2005

FOSTER, Nancy F. ; GIBBONS, Susan: Understanding Faculty to Improve Content Recruitment for Institutional Repositories. In: *D-Lib Magazine* 11 (2005), Nr. 1. <<http://dx.doi.org/10.1045/january2005-foster>>, Abruf: 10.08.2011

### Foulonneau u. André 2008

FOULONNEAU, Muriel ; ANDRÉ, Francis: *Investigative Study of Standards for Digital Repositories and Related Services*. Amsterdam : Amsterdam University Press, 2008 <<http://dare.uva.nl/document/93727>>, Abruf: 10.08.2011

### Foulonneau u. Riley 2008

FOULONNEAU, Muriel ; RILEY, Jenn.: *Metadata for digital resources: implementation, systems design and interoperability*. Oxford : Chandos, 2008 (Chandos information professional series)

### Fry u. a. 2009

FRY, Jenny ; OPPENHEIM, Charles ; PROBETS, Steve ; CREASER, Claire ; GREENWOOD, Helen ; SPEZI, Valérie ; WHITE, Sonya ; LISU (Hrsg.) ; LOUGHSBOROUGH UNIV. (Hrsg.): *PEER Behavioural Research: Authors and Users vis-à-vis Journals and Repositories; Baseline Report*. 2009 <[http://www.peerproject.eu/fileadmin/media/reports/Final\\_revision\\_-\\_behavioural\\_baseline\\_report\\_-\\_20\\_01\\_10.pdf](http://www.peerproject.eu/fileadmin/media/reports/Final_revision_-_behavioural_baseline_report_-_20_01_10.pdf)>, Abruf: 10.08.2011

**Gantert u. Hacker 2008**

GANTERT, Klaus ; HACKER, Rupert: *Bibliothekarisches Grundwissen*. 8. Aufl. München : Saur, 2008

**Goescholar 2011**

GOESCHOLAR: *Hilfe und FAQ*. <<http://goedoc.uni-goettingen.de/goescholar/help>>. Version: 2011, Abruf: 10.08.2011

**Gold 2007**

GOLD, Anna: Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians. In: *D-Lib Magazine* 13 (2007), Nr. 9/10. <<http://www.dlib.org/dlib/september07/gold/09gold-pt1.html>>, Abruf: 10.08.2011

**Graaf 2009**

GRAAF, Maurits van der: *The European Repository Landscape 2008: Inventory of Digital Repositories for Research Output*. Amsterdam : Amsterdam Univ. Press, 2009 <<http://dare.uva.nl/aup/nl/record/316871>>, Abruf: 10.08.2011

**Green u. Awre 2007**

GREEN, Richard ; AWRE, Chris: *RepoMMan Project: Automatic Generation of Object Metadata*. 2007 <<http://www.hull.ac.uk/esig/repomman/downloads/D-D13-autometadata-v11.pdf>>, Abruf: 10.08.2011

**Grötschel u. a. 1993**

GRÖTSCHEL, Martin ; LÜGGER, Joachim ; SPERBER, Wolfram: Wissenschaftliches Publizieren und Elektronische Fachinformation im Umbruch: ein Situationsbericht aus der Sicht der Mathematik / Konrad-Zuse-Zentrum für Informationstechnik Berlin. Version: 1993. <<http://opus4.kobv.de/opus4-zib/files/496/TR-93-04.pdf>>, Abruf: 10.08.2011. 1993 (TR-93-04). – Forschungsbericht

**Guy u. a. 2004**

GUY, Marieke ; POWELL, Andy ; DAY, Michael: Improving the Quality of Metadata in Eprint Archives. In: *Ariadne* 38 (2004), January. <<http://www.ariadne.ac.uk/issue38/guy/>>, Abruf: 10.08.2011

**Hagenhoff u. a. 2007**

HAGENHOFF, Svenja ; SEIDENFADEN, Lutz ; ORTELBACH, Björn ; SCHUMANN, Matthias: *Göttinger Schriften zur Internetforschung*. Bd. 4: *Neue Formen der Wissenschaftskommunikation: Eine Fallstudienuntersuchung*. Göttingen : Universitäts-

verl., 2007 <[www2.as.wiwi.uni-goettingen.de/getfile?DateiID=642](http://www2.as.wiwi.uni-goettingen.de/getfile?DateiID=642)>, Abruf: 10.08.2011

#### **Haller 1998**

HALLER, Klaus: *Katalogkunde : eine Einführung in die Formal- und Sacherschließung*. 3., erw. Aufl. München : Saur, 1998

#### **Haller u. Fabian 2004**

HALLER, Klaus ; FABIAN, Claudia: Bestandserschließung. In: RUDOLF FRANKENBERGER AND KLAUS HALLER (Hrsg.): *Die moderne Bibliothek : ein Kompendium der Bibliotheksverwaltung*. München : Saur, 2004, 222–261

#### **Harnad 1995**

HARNAD, Stevan: *Overture: The Subversive Proposal [1994]*. <<http://www.arl.org/sc/subversive/i-overture-the-subversive-proposal.shtml>>. Version: 1995, Abruf: 10.08.2011

#### **Herb 2009**

HERB, Ulrich: *Chancen im OPUS: Automatisiert SWD-Schlagwörter produzieren*. 2009 <<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bsz:ch1-200901378>>, Abruf: 10.08.2011. – (33. Jahrestagung der Gesellschaft für Klassifikation)

#### **Hillmann 2005**

HILLMANN, Diane: *Using Dublin Core - The Elements*. <<http://dublincore.org/documents/2005/11/07/usageguide/elements.shtml>>. Version: November 2005, Abruf: 10.08.2011

#### **Hillmann u. a. 2004**

HILLMANN, Diane ; DUSHAY, Naomi ; PHIPPS, Jon: Improving metadata quality: augmentation and recombination. In: *Proceedings of the 2004 international conference on Dublin Core and metadata applications: metadata across languages and cultures*, Dublin Core Metadata Initiative, 2004, 7:1–7:8

#### **Hogenaar 2009**

HOGENAAR, Arjan: Enhancing Scientific Communication through Aggregated Publications Environments. In: *Ariadne* (2009), Nr. 61. <<http://www.ariadne.ac.uk/issue61/hogenaar/>>, Abruf: 10.08.2011

## **IFLA 2004**

IFLA: *IFLA Statement on Open Access to Scholarly Literature and Research Documentation*. <<http://www.ifla.org/V/cdoc/open-access04.html>>. Version: Dezember 2004, Abruf: 10.08.2011

## **IFLA Study Group on the Functional Requirements for Bibliographic Records 2008**

IFLA STUDY GROUP ON THE FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS: *Functional Requirements for Bibliographic Records*. 2008 <[http://www.ifla.org/files/cataloguing/frbr/frbr\\_2008.pdf](http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf)>, Abruf: 10.08.2011

## **Joint 2006**

JOINT, Nicholas: Institutional Repositories, Self-Archiving and the Role of the Library. In: *Library Review* 55 (2006), Nr. 2, 81–84. <<http://dx.doi.org/10.1108/00242530610649576>>, Abruf: 10.08.2011

## **Kindling u. a. 2010**

KINDLING, Maxi ; MÜLLER, Uwe ; PAMPEL, Heinz: *Thesen zur Zukunft von Open-Access-Repositorien*. <<http://www.dini.de/veranstaltungen/workshops/dinihelmholtz-workshop-repositorien-praxis-und-vision/abstracts/podiumsdiskussion/>>. Version: November 2010, Abruf: 10.08.2011

## **Kirchgässner 2003**

KIRCHGÄSSNER, Adalbert: *13 Jahre Zeitschriftenabbestellung an der Universität Konstanz*. 2003 <<http://kops.ub.uni-konstanz.de/volltexte/2003/1036>>, Abruf: 10.08.2011. – (Vortrag, gehalten auf der Diskussionsveranstaltung 'Das Zeitschriftenparadoxon oder wer verfügt über wissenschaftliche Information' am 20. Mai 2003 in Stuttgart)

## **KOPS 2011**

KOPS: *Hilfe zu KOPS: Publikation anmelden*. <<http://kops.ub.uni-konstanz.de/help>>. Version: 2011, Abruf: 09.08.2011

## **Krönung u. a. 2010**

KRÖNUNG, Julia ; BERNIUS, Steffen ; BOSCH, Kai-G. ; DUGALL, Berndt: Durch Selbstarchivierung und Nationallizenzen zu Open Access? In: *ABI-Technik* 30 (2010), Nr. 4, 230–239. <<http://www.wi-frankfurt.de/publikationenNeu/DurchSelbstarchivierungundNati4066.pdf>>, Abruf: 10.08.2011



### **Kölbl 2011**

KÖLBL, Matthias: *Wissensmanagement in der Wissenschaft: das deutsche Wissenschaftssystem und sein Beitrag zur Bewältigung gesellschaftlicher Herausforderungen*. 2. elektronische Auflage 2011. Berlin : Gesellschaft für Wissenschaftsforschung, 2011 <<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:101:1-201107213687>>, Abruf: 10.08.2011

### **Lagoze u. Sompel 2008**

LAGOZE, Carl (Hrsg.) ; SOMPEL, Herbert van d. (Hrsg.): *The Open Archives Initiative Protocol for Metadata Harvesting*. 2008 <<http://www.openarchives.org/OAI/openarchivesprotocol.htm>>, Abruf: 10.08.2011

### **Lynch 2003**

LYNCH, Clifford A.: Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. In: *ARL* (2003), Februar, Nr. 226, 1–7. <<http://www.arl.org/resources/pubs/br/br226/br226ir.shtml>>, Abruf: 10.08.2011

### **Mapping SWAP 2008**

REPOSITORIES RESEARCH TEAM (JISC) (Hrsg.): *Mapping the Scholarly Works Application Profile to Simple DC*. <[http://www.ukoln.ac.uk/repositories/digirep/index/Mapping\\_the\\_Scholarly\\_Works\\_Application\\_Profile\\_to\\_Simple\\_DC](http://www.ukoln.ac.uk/repositories/digirep/index/Mapping_the_Scholarly_Works_Application_Profile_to_Simple_DC)>. Version: Mai 2008, Abruf: 10.08.2011. – (DigiRepWiki)

### **Mayr u. Petras 2008**

MAYR, Philipp ; PETRAS, Vivien: *Crosskonkordanzen : Terminologiemapping und deren Effektivität für das Information Retrieval*. 2008 (Vortrag World Library and Information Congress: 74th IFLA General Conference and Council). <[http://archive.ifla.org/IV/ifla74/papers/129-Mayr\\_Petras-trans-de.pdf](http://archive.ifla.org/IV/ifla74/papers/129-Mayr_Petras-trans-de.pdf)>, Abruf: 05.08.2010

### **Mazzoni 2009**

MAZZONI, Roberto: *Abschluss Projekt Open Access*. <<http://www.id.uzh.ch/zinfo/openaccess.html>>. Version: 2009, Abruf: 02.08.2011 (ZInfo (Newsletter der Informatikdienste der Univ. Zürich)

### **McKay 2007**

MCKAY, Dana: Institutional Repositories and Their 'Other' Users: Usability Beyond Authors. In: *Ariadne* (2007), Nr. 52. <<http://www.ariadne.ac.uk/issue52/mckay/>>, Abruf: 10.08.2011



### **Mingam 2005**

MINGAM, Michel: Rameau: bilan, perspectives. In: *Bulletin des bibliothèques de France* 50 (2005), Nr. 5, 38–47. <<http://bbf.enssib.fr/consulter/bbf-2005-05-0038-001.pdf>>, Abruf: 10.08.2011

### **Moen u. a. 1997**

MOEN, William. E. ; STEWART, Erin L. ; MCCLURE, Charles R.: *The Role of Content Analysis in Evaluating Metadata for the U.S. Government Information Locator Service (GILS): Results from an Exploratory Study*. <<http://www.unt.edu/wmoen/publications/GILSMDContentAnalysis.htm>>. Version: 1997, Abruf: 10.08.2011

### **Müller u. a. 2009**

MÜLLER, Uwe ; SEVERIENS, Thomas ; MALITZ, Robin ; SCHIRMBACHER, Peter: OA Network: An Integrative Open Access Infrastructure for Germany . In: *D-Lib Magazine* 15 (2009), Nr. 9/10. <<http://dx.doi.org/10.1045/september2009-mueller>>, Abruf: 10.08.2011

### **Neuroth 2009**

NEUROTH, Heike: *eScience und Bibliotheken*. 2009 <[http://www.fbi.fh-koeln.de/symposium/vortraege/FHKoeln\\_Neuroth.pdf](http://www.fbi.fh-koeln.de/symposium/vortraege/FHKoeln_Neuroth.pdf)>, Abruf: 10.08.2011. – (Präsentation auf dem Symposium der FH Köln „Von den Preußischen Instruktionen zu eScience“, 29.01.2009)

### **Open Access Directory 2010**

OPEN ACCESS DIRECTORY: *Timeline of the Open Access Movement: Timeline Before 2000*. <[http://oad.simmons.edu/oadwiki/Timeline\\_before\\_2000](http://oad.simmons.edu/oadwiki/Timeline_before_2000)>. Version: November 2010, Abruf: 10.08.2011 (Open Access Directory Wiki)

### **Oppenheim 2008**

OPPENHEIM, Charles: Electronic scholarly publishing and open access. In: *Journal of Information Science* 34 (2008), Nr. 4, 577–590. <<http://dx.doi.org/10.1177/0165551508092268>>, Abruf: 10.08.2011

### **OPUS 2011**

OPUS: *Liste der produktiven OPUS-Installationen*. <<http://www.opus-repository.org/anwendung/installationen.html>>. Version: 2011, Abruf: 10.08.2011

**Orsdel u. Born 2009**

ORSDEL, Lee C. V. ; BORN, Kathleen: Reality Bites: Periodicals Price Survey 2009 . In: *Library Journal* 134 (2009), Nr. 7, 36–40. – (Online als HTML-Version verfügbar unter: <<http://www.libraryjournal.com/article/CA6651248.html>> [Abruf: 10.08.2011])

**Oßwald 2008**

OSSWALD, Achim: E-science and information services: a missing link in the context of digital libraries. In: *Online Information Review* 32 (2008), Nr. 4, 516–523. <<http://dx.doi.org/10.1108/14684520810897395>>, Abruf: 10.08.2011

**Park u. Tosaka 2010**

PARK, Jung-Ran ; TOSAKA, Yuji: Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms. In: *Cataloging & Classification Quarterly* 48 (2010), November, Nr. 8, 696–715. <<http://dx.doi.org/10.1080/01639374.2010.508711>>, Abruf: 10.08.2011. – ISSN 0163–9374

**Pieper u. Wolf 2009**

PIEPER, Dirk ; WOLF, Sebastian: Wissenschaftliche Dokumente in Suchmaschinen. Version: 2009. <<http://hdl.handle.net/10760/12746>>, Abruf: 10.08.2011. In: DIRK, Lewandowski (Hrsg.): *Handbuch Internet-Suchmaschinen: Nutzerorientierung in Wissenschaft und Praxis*. Heidelberg : Akad. Verl.-Gesellschaft, 2009, 356–374

**Plassmann u. a. 2006**

PLASSMANN, Engelbert ; RÖSCH, Hermann ; SEEFELD, Jürgen ; UMLAUF, Konrad: *Bibliotheken und Informationsgesellschaft in Deutschland*. Wiesbaden : Harrassowitz, 2006

**Reimer 2011**

REIMER, Torsten: *Virtuelle Forschungsumgebungen im internationalen Kontext – Die neue (Digitale) Bibliothek?* 2011 <<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:bsz:nn-opus-10218>>, Abruf: 10.08.2011. – Präsentation auf dem 100. Deutscher Bibliothekartag

**Research Information Network 2010**

RESEARCH INFORMATION NETWORK (Hrsg.): *Research Support Services in UK Universities*. 2010 <[http://www.rin.ac.uk/system/files/attachments/Research\\_Support\\_Services\\_in\\_UK\\_Universities\\_report\\_for\\_screen.pdf](http://www.rin.ac.uk/system/files/attachments/Research_Support_Services_in_UK_Universities_report_for_screen.pdf)>, Abruf: 10.08.2011

### **Robertson u. Robertson 2006**

ROBERTSON, Suzanne ; ROBERTSON, James: *Mastering the requirements process*. 2. ed. Upper Saddle River, NJ : Addison-Wesley, 2006

### **Salo 2008**

SALO, Dorothea: Innkeeper at the Roach Motel. In: *Library Trends* 57 (2008), Nr. 2, 98–123. <<http://hdl.handle.net/2142/10678>>, Abruf: 10.08.2011

### **Salo 2009**

SALO, Dorothea: Name Authority Control in Institutional Repositories. In: *Cataloging & Classification Quarterly* 47 (2009), Nr. 3, 249–261. <<http://dx.doi.org/10.1080/01639370902737232>>, Abruf: 10.08.2011. – ISSN 0163–9374

### **Schöning-Walther 2011**

SCHÖNING-WALTHER, Christa: Automatische Erschließungsverfahren für Netzpublikationen. In: *Dialog mit Bibliotheken* (2011), Nr. 1, 3–36. <[http://files.d-nb.de/pdf/petrus/petrus\\_dialog\\_2011\\_1.pdf](http://files.d-nb.de/pdf/petrus/petrus_dialog_2011_1.pdf)>, Abruf: 10.08.2011

### **SHERPA RoMEO 2006**

SHERPA ROMEO: *Colours, Pre-print, Post-print, Definitions and Terms*. <<http://www.sherpa.ac.uk/romeoinfo.html>>. Version: 2006, Abruf: 10.08.2011

### **Shreeves u. a. 2005**

SHREEVES, Sarah L. ; KNUTSON, Ellen M. ; STVILIA, Besiki ; PALMER, Carole L. ; TWIDALE, Michael B. ; COLE, Timothy W.: Is ‘Quality’ Metadata ‘Shareable’ Metadata? The Implications of Local Metadata Practice on Federated Collections. Version: 2005. <<http://hdl.handle.net/2142/145>>. In: THOMPSON, Hugh A. (Hrsg.): *Proceedings of the Twelfth National Conference of the Association of College and Research Libraries, April 7-10 2005, Minneapolis, MN*. Chicago : Association of College and Research Libraries, lastchecked = 10.08.2011, 2005, 223–237

### **Springer 2011a**

SPRINGER: *Quality & Quantity : Instructions for Authors*. <[http://www.springer.com/social+sciences/journal/11135?print\\_view=true&detailsPage=pltc\\_i\\_1229508](http://www.springer.com/social+sciences/journal/11135?print_view=true&detailsPage=pltc_i_1229508)>. Version: 2011, Abruf: 08.08.2011

### **Springer 2011b**

SPRINGER: *Springer’s Self-Archiving Policy*. <<http://www.springer.com/open+>

[access/authors+rights?SGWID=0-176704-12-683201-0](#)>. Version: 2011, Abruf: 01.08.2011

#### **Stvilia u. a. 2004**

STVILIA, Besiki ; GASSER, Les ; TWIDALE, Michael B. ; SHREEVES, Sarah L. ; COLE, Tim W.: Metadata quality for federated collections. Version: 2004, lastchecked = 10.08.2011. <<http://hdl.handle.net/2142/721>>. In: CHENGULUR-SMITH, S. (Hrsg.) ; RASCHID, L. (Hrsg.) ; LONG, J. (Hrsg.) ; SEKO, C. (Hrsg.): *Proceedings of ICIQ04-9th International Conference on Information Quality*. Cambridge, MA : MIT, 2004, lastchecked = 10.08.2011, 111–125

#### **SUB Göttingen 2011**

SUB GÖTTINGEN: *Open Access-Zeitschriften: Optionen für Wissenschaftler*. <[http://www.sub.uni-goettingen.de/ebene\\_1/1\\_oa\\_journals.html.de](http://www.sub.uni-goettingen.de/ebene_1/1_oa_journals.html.de)>. Version: 2011, Abruf: 09.08.2011

#### **SWAP 2009**

REPOSITORIES RESEARCH TEAM (JISC) (Hrsg.): *Scholarly Works Application Profile*. <[http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly\\_Works\\_Application\\_Profile](http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile)>. Version: März 2009, Abruf: 10.08.2011. – (Digi-RepWiki)

#### **Taylor u. Joudrey 2009**

TAYLOR, Arlene G. ; JOUDREY, Daniel N.: *The organization of information*. 3. Aufl. Westport, Conn. [u.a.] : Libraries Unlimited, 2009 (Library and information science text series)

#### **Univ. Gießen 2005**

*Promotionsordnung vom 4. Februar 2005 der Gemeinsamen Kommission Naturwissenschaften für die Naturwissenschaftlichen Fachbereiche der Justus-Liebig-Universität Gießen*. 2005 <[http://www.uni-giessen.de/cms/fbz/fb07/Downloads/Rechtsgrundlagen/promotionsordnung-4-2-2005.pdf/at\\_download/file](http://www.uni-giessen.de/cms/fbz/fb07/Downloads/Rechtsgrundlagen/promotionsordnung-4-2-2005.pdf/at_download/file)>, Abruf: 10.08.2011

#### **Univ. Witten 2005**

*Promotionsordnung vom 4. Februar 2005*. 2005 <[http://www.uni-wh.de/fileadmin/media/w/w\\_studium/w\\_promotion\\_habilitation/WiWi-Prom-0-Ausführungsbestimmungen.pdf](http://www.uni-wh.de/fileadmin/media/w/w_studium/w_promotion_habilitation/WiWi-Prom-0-Ausführungsbestimmungen.pdf)>, Abruf: 10.08.2011

### **Univ. Zürich 2009**

UNIV. ZÜRICH: *OA an der UZH*. <<http://www.oai.uzh.ch/index.php?option=content&task=view&id=368&Itemid=246>>. Version: 2009, Abruf: 10.08.2011

### **Univ. Zürich 2010**

*Doktoratsordnung der Allgemeinen Doktoratsstufe der Philosophischen Fakultät der Universität Zürich*. 2010 <[http://www.phil.uzh.ch/studium/studium/downloads/studienordnung/doktorat/\\_Doktoratsordnungen\\_PhF\\_AllgemeinesDoktorat.pdf](http://www.phil.uzh.ch/studium/studium/downloads/studienordnung/doktorat/_Doktoratsordnungen_PhF_AllgemeinesDoktorat.pdf)>, Abruf: 10.08.2011

### **Univ. Zürich 2011a**

UNIV. ZÜRICH, Hauptbibliothek: *Anleitung: Eingabe von Publikationen in Zurich Open Repository and Archive (ZORA)*. 2011 <[http://www.oai.uzh.ch/images/stories/eingabe\\_in\\_zora\\_anleitung/ZORA\\_Eingabe\\_2011\\_05\\_30.pdf](http://www.oai.uzh.ch/images/stories/eingabe_in_zora_anleitung/ZORA_Eingabe_2011_05_30.pdf)>, Abruf: 10.08.2011

### **Univ. Zürich 2011b**

UNIV. ZÜRICH, Hauptbibliothek: *Leitfaden zu ZORA*. <<http://www.oai.uzh.ch/index.php?option=content&task=view&id=410&Itemid=286>>. Version: Mai 2011, Abruf: 10.08.2011

### **Univ. Zürich 2011c**

UNIV. ZÜRICH, Hauptbibliothek: *Über ZORA: Import*. <[http://www.oai.uzh.ch/index.php?option=content&task=view&id=436&Itemid=312&mos\\_lng=de](http://www.oai.uzh.ch/index.php?option=content&task=view&id=436&Itemid=312&mos_lng=de)>. Version: März 2011, Abruf: 10.08.2011

### **US National Institutes of Health 2010**

US NATIONAL INSTITUTES OF HEALTH: *US National Institutes of Health Public Access*. <<http://publicaccess.nih.gov>>. Version: 2010, Abruf: 10.08.2011

### **Vaart 2010**

VAART, Lilian van der: *Collaboratories: Connecting Researchers. How to facilitate choice, design and uptake of online research collaboratories*. Utrecht : SURF, 2010 <<http://www.surffoundation.nl/SiteCollectionDocuments/Collaboratories%20Connecting%20Researchers.pdf>>, Abruf: 10.08.2011

### **Vanderfeesten u. a. 2008**

VANDERFEESTEN, Maurice (Hrsg.) ; SUMMANN, Friedrich (Hrsg.) ; SLABBERTJE, Martin (Hrsg.): *DRIVER Guidelines 2.0 : Guidelines for Content*

*Providers - Exposing Textual Resources with OAI-PMH.* DRIVER, 2008  
<[http://www.driver-support.eu/documents/DRIVER\\_Guidelines\\_v2\\_Final\\_2008-11-13.pdf](http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf)>, Abruf: 10.08.2011

#### **Warner 2010**

WARNER, Simeon: Author Identifiers in Scholarly Repositories. In: *Journal of Digital Information* 11 (2010), Nr. 1. <<http://journals.tdl.org/jodi/article/view/752/641>>, Abruf: 10.08.2011

#### **Weller u. a. 2007**

WELLER, Katrin ; MAINZ, Indra ; PAULSEN, Ingo ; MAINZ, Dominic: Semantisches und vernetztes Wissensmanagement für Forschung und Wissenschaft. Version: 2007. <[http://epub.uni-regensburg.de/4917/1/Ball\\_2007.Wisskom.pdf](http://epub.uni-regensburg.de/4917/1/Ball_2007.Wisskom.pdf)>, Abruf: 10.08.2011. In: BALL, Rafael (Hrsg.): *Wissenschaftskommunikation der Zukunft : WissKom 2007, 4. Konferenz der Zentralbibliothek Forschungszentrum Jülich*. Jülich : Forschungszentrum Jülich, 2007 (Schriften des Forschungszentrums Jülich : Reihe Bibliothek / Library ; 18), 33–46

#### **White u. Creaser 2007**

WHITE, Sonya ; CREASER, Claire: *LISU Occasional Paper*. Bd. 37: *Trends in Scholarly Journal Prices 2000–2006*. Loughsbrough : LISU, 2007 <<http://www.lboro.ac.uk/departments/ls/lisu/downloads/op37.pdf>>, Abruf: 10.08.2011

#### **Wiesner 2004**

WIESNER, Margot: *Global Journal Publishing: Different Markets – Common Strategies?* 2004 <[http://www.hebis.de/hebis-konsortium/presentationen/Wiesner\\_Orlando\\_text.pdf](http://www.hebis.de/hebis-konsortium/presentationen/Wiesner_Orlando_text.pdf)>, Abruf: 10.08.2011. – Vortrag, 2004 ALA Annual Conference, Orlando, FL.

#### **Wiesner u. a. 2004**

WIESNER, Margot ; WERNER, Andreas ; SCHÄFFLER, Hildegard: Bestandsaufbau. In: RUDOLF FRANKENBERGER AND KLAUS HALLER (Hrsg.): *Die moderne Bibliothek : ein Kompendium der Bibliotheksverwaltung*. München : Saur, 2004, 166–221

#### **Wissenschaftsrat 2001**

WISSENSCHAFTSRAT: *Empfehlungen zur digitalen Informationsversorgung durch Hochschulbibliotheken (Drs. 4935/01)*. 2001 <<http://www.wissenschaftsrat.de/download/archiv/4935-01.pdf>>, Abruf: 10.08.2011

### **Zeitschriftendatenbank 2011a**

ZEITSCHRIFTENDATENBANK: *Journals Online Print.* <<http://www.zeitschriftendatenbank.de/services/journals-online-print.html>>.

Version: Juli 2011, Abruf: 10.08.2011

### **Zeitschriftendatenbank 2011b**

ZEITSCHRIFTENDATENBANK: *SRU.* <<http://www.zeitschriftendatenbank.de/services/schnittstellen/sru.html>>. Version: September 2011, Abruf:

09.08.2011

# A Verwendete Datensätze

## A.1 Datensatz *Goescholar*

Beispielsatz Goescholar (Datenformat: Qualified DC):

<http://goedoc.uni-goettingen.de/oai/request?verb=GetRecord&metadataPrefix=qdc&identifier=oai%3Ahttp%3A%2F%2Fgoedoc.uni-goettingen.de%3Agoescholar%2F2089>

```
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
<responseDate>2011-08-10T17:40:53Z</responseDate>
<request metadataPrefix="qdc" identifier="oai:http://goedoc.uni-goettingen.de:goescholar/1270"
verb="GetRecord">http://goedoc.uni-goettingen.de/oai/request</request>
<GetRecord>
<record>
<header>
<identifier>
oai:http://goedoc.uni-goettingen.de:goescholar/1270
</identifier>
<timestamp>2011-03-07T15:03:07Z</timestamp>
<setSpec>hdl_goescholar_2667</setSpec>
</header>
<metadata>
<dc:creator xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">Mohammed, Yassene</dc:creator>
<dc:creator xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd" xml:lang="eng">
Verhey, Janko F.</dc:creator>
<dc:identifier type="dcterms:URI" xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd" xml:lang="eng">
http://resolver.sub.uni-goettingen.de/purl?goescholar/1270
</dc:identifier>
<dcterms:abstract xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd" xml:lang="eng">
Background: Laser Interstitial ThermoTherapy (LITT) is a well established
```



surgical method. The use of LITT is so far limited to homogeneous tissues, e.g. the liver. One of the reasons is the limited capability of existing treatment planning models to calculate accurately the damage zone. The treatment planning in inhomogeneous tissues, especially of regions near main vessels, poses still a challenge. In order to extend the application of LITT to a wider range of anatomical regions new simulation methods are needed. The model described with this article enables efficient simulation for predicting damaged tissue as a basis for a future laser-surgical planning system. Previously we described the dependency of the model on geometry. With the presented paper including two video files we focus on the methodological, physical and mathematical background of the model. Methods: In contrast to previous simulation attempts, our model is based on finite element method (FEM). We propose the use of LITT, in sensitive areas such as the neck region to treat tumours in lymph node with dimensions of 0.5 cm 2 cm in diameter near the carotid artery. Our model is based on calculations describing the light distribution using the diffusion approximation of the transport theory; the temperature rise using the bioheat equation, including the effect of microperfusion in tissue to determine the extent of thermal damage; and the dependency of thermal and optical properties on the temperature and the injury. Injury is estimated using a damage integral. To check our model we performed a first in vitro experiment on porcine muscle tissue.

Results: We performed the derivation of the geometry from 3D ultrasound data and show for this proposed geometry the energy distribution, the heat elevation, and the damage zone. Further on, we perform a comparison with the in-vitro experiment. The calculation shows an error of 5% in the x-axis parallel to the blood vessel.

</dcterms:abstract>

<dc:format type="dcterms:IMT" xsi:schemaLocation="http://purl.org/dc/terms/http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd http://purl.org/dc/elements/1.1/http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd" xml:lang="eng">application/pdf</dc:format>  
 <dc:language type="dcterms:ISO639-3" xsi:schemaLocation="http://purl.org/dc/terms/http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd http://purl.org/dc/elements/1.1/http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd" xml:lang="eng">eng</dc:language>  
 <dcterms:bibliographicCitation xsi:schemaLocation="http://purl.org/dc/terms/http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd http://purl.org/dc/elements/1.1/http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"

```

xml:lang="eng">
Biomedical engineering online; Vol. 4, Nr. 2, p. 1-16
</dcterms:bibliographicCitation>
<dc:subject xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd http://purl.org/dc/elements/1.1/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd" xml:lang="eng">
Medizinische Mathematik</dc:subject>
<dc:subject schema="DDC" xsi:schemaLocation=
"http://purl.org/dc/terms/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">004</dc:subject>
<dc:subject schema="DDC" xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">610</dc:subject>
<dc:title xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">
A finite element method model to simulate laser interstitial thermo therapy in
anatomical inhomogeneous regions
</dc:title>
<dc:subject schema="BK" xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">44.32</dc:subject>
<dcterms:isPartOf xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">Biomedical engineering online</dcterms:isPartOf>
<dcterms:issued xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">2005</dcterms:issued>
<dc:type xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">info:eu-repo/semantics/article</dc:type>
<dc:type xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">info:eu-repo/semantics/publishedVersion</dc:type>
<dc:rights type="dcterms:URI" xsi:schemaLocation="http://purl.org/dc/terms/

```

```

http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">info:eu-repo/semantics/openAccess</dc:rights>
<dc:rights type="dcterms:URI" xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="deu">http://goedoc.uni-goettingen.de/goeschloar/rights</dc:rights>
<dc:identifier type="dcterms:URI" xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">http://dx.doi.org/10.1186/1475-925X-4-2</dc:identifier>
<dc:identifier type="dcterms:URI" xsi:schemaLocation="http://purl.org/dc/terms/
http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd
http://purl.org/dc/elements/1.1/ http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
xml:lang="eng">info:ppn:526362898</dc:identifier>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

## A.2 Datensatz *BiPrints*

### Beispielsatz Biprints (Datenformat: DC Simple)

[http://repositories.ub.uni-bielefeld.de/biprints/oai2/oai2.php?verb=GetRecord&metadataPrefix=oai\\_dc&identifier=oai:biprints.ub.uni-bielefeld.de:4583](http://repositories.ub.uni-bielefeld.de/biprints/oai2/oai2.php?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:biprints.ub.uni-bielefeld.de:4583)

```

<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="oai.xsl" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/
2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
<responseDate>2011-07-31T14:14:58Z</responseDate>
<request verb="GetRecord" metadataPrefix="oai_dc"
identifier="oai:biprints.ub.uni-
bielefeld.de:4583">http://repositories.ub.uni-
bielefeld.de/biprints/oai2/oai2.php</request>
<GetRecord>
<record>
<header>
<identifier>oai:biprints.ub.uni-

```

```

bielefeld.de:4583</identifier>
<timestamp>2010-07-05T15:21:21Z</timestamp>
<setSpec>ddc:610</setSpec>
<setSpec>pub-type:2</setSpec>
</header>
<metadata>
<oai_dc:dc
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_
dc/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.openarchives.org/OAI/2.
0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>&#039;Imported risk&#039; or &#039;health
transition&#039;? Smoking prevalence among ethnic German
immigrants from the Former Soviet Union by duration of stay
in Germany - analysis of microcensus data</dc:title>
<dc:creator>Reiss, Katharina</dc:creator>
<dc:creator>Spallek, Jacob</dc:creator>
<dc:creator>Razum, Oliver</dc:creator>
<dc:subject>Medicine and health</dc:subject>
<dc:description>Background
It can be assumed that resettlers (ethnic German immigrants
from the Former Soviet Union) show similar smoking patterns
as persons in their countries of origin at the time of
migration. We analysed how the smoking prevalence among
resettlers differs from that among the general population of
Germany and whether the prevalence differs between groups
with increasing duration of stay.
Methods
To estimate the smoking prevalence we used the scientific-
use-file (n = 477,239) of the German 2005 microcensus, an
annual census representing 1% of all German households.
Participation in the microcensus is obligatory (unit-
nonresponse <7%). We stratified the prevalence of smoking
among resettlers and the comparison group (population of
Germany without resettlers) by age, sex, educational level
and duration of stay. In total, 14,373 (3% of the total)
persons were identified as resettlers.
Results
Female resettlers with short duration of stay had a
significantly lower smoking prevalence than women in the

```

comparison group. With increasing duration of stay their smoking prevalence appears to converge to that of the comparison group (e.g.: high educational level, age group 25-44 years: short duration of stay 15%, long duration of stay 24%, comparison group 28%). In contrast, the smoking prevalence among male resettlers with short duration of stay was significantly higher than that among men in the comparison group, but also with a trend towards converging (e.g.: high educational level, age group 25-44 years: short duration of stay 44%, long duration of stay 35%, comparison group 36%). Except for female resettlers with short duration of stay, the participants with low educational level had on average a higher smoking prevalence than those with a high educational level.

#### Conclusions

This is the first study estimating the smoking prevalence among resettlers by duration of stay. The results support the hypothesis that resettlers brought different smoking habits from their countries of origin shortly after migration. The observed convergence of the smoking habits with increasing duration of stay is in line with the hypothesis of migration as 'health transition'. However, due to the cross-sectional design of the study, further research is needed to confirm these findings.

<dc:publisher>Universität Bielefeld</dc:publisher>

<dc:publisher>Fakultäten. Fakultät für  
Gesundheitswissenschaften</dc:publisher>

<dc:date>2010</dc:date>

<dc:type>Journal Article</dc:type>

<dc:format>application/pdf</dc:format>

<dc:identifier>http://repositories.ub.uni-  
bielefeld.de/biprints/volltexte/2010/4583/</dc:identifier>

<dc:source>International Journal for Equity in Health,  
9:15</dc:source>

<dc:language>eng</dc:language>

<dc:rights>Metadata supplied by: Universität Bielefeld,  
Germany</dc:rights>

</oai\_dc:dc>

</metadata>

</record>

</GetRecord>

</OAI-PMH>

## A.3 Datensatz *KOPS*

### Beispielsatz *KOPS* (Datenformat *METS*)

<http://kops.ub.uni-konstanz.de/oai/request?verb=GetRecord&metadataPrefix=mets&identifier=oai:kops.ub.uni-konstanz.de:urn:nbn:de:bsz:352-opus-102908>

```
<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
<responseDate>2011-07-31T12:44:42Z</responseDate>
<request identifier="oai:kops.ub.uni-konstanz.de:urn:nbn:de:bsz:352-opus-102908"
metadataPrefix="mets" verb="GetRecord">http://kops.ub.uni-
konstanz.de/oai/request</request>
<GetRecord>
<record>
<header>
<identifier>
oai:kops.ub.uni-konstanz.de:urn:nbn:de:bsz:352-opus-102908
</identifier>
<datestamp>2011-03-24T18:14:12Z</datestamp>
<setSpec>coll-69</setSpec>
</header>
<metadata>
<mets OBJID="hdl:urn:nbn:de:bsz:352-opus-102908" LABEL="DSpace Item"
xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
<metsHdr CREATEDATE="2011-07-31T14:44:42">
<agent ROLE="CUSTODIAN" TYPE="ORGANIZATION">
<name>DSpace at My University</name>
</agent>
</metsHdr>
<dmdSec ID="DMD_urn_nbn_de_bsz_352-opus-102908">
<mdWrap MDTYPE="MODS">
<xmlData>
<mods:name>
<mods:role>
<mods:roleTerm type="text">author</mods:roleTerm>
</mods:role>
<mods:namePart>Holzberger, Bastian</mods:namePart>
</mods:name>
<mods:name>
<mods:role>
<mods:roleTerm type="text">author</mods:roleTerm>
```

```

</mods:role>
<mods:namePart>Rubini, Marina</mods:namePart>
</mods:name>
<mods:name>
<mods:role>
<mods:roleTerm type="text">author</mods:roleTerm>
</mods:role>
<mods:namePart>Möller, Heiko M.</mods:namePart>
</mods:name>
<mods:name>
<mods:role>
<mods:roleTerm type="text">author</mods:roleTerm>
</mods:role>
<mods:namePart>Marx, Andreas</mods:namePart>
</mods:name>
<mods:extension>
<mods:dateAccessioned encoding="iso8601">2011-03-
24T18:14:12Z</mods:dateAccessioned>
</mods:extension>
<mods:originInfo>
<mods:dateIssued encoding="iso8601">2010</mods:dateIssued>
</mods:originInfo>
<mods:relatedItem type="host">
<mods:part>
<mods:text>
First publ. in: Angewandte Chemie International Edition 49 (2010), 7, pp. 1324-1327
</mods:text>
</mods:part>
</mods:relatedItem>
<mods:identifier type="uri">
http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-102908
</mods:identifier>
<mods:physicalDescription>
<mods:internetMediaType>application/pdf</mods:internetMediaType>
</mods:physicalDescription>
<mods:language>
<mods:languageTerm authority="rfc3066">eng</mods:languageTerm>
</mods:language>
<mods:accessCondition type="useAndReproduction">deposit-
license</mods:accessCondition>
<mods:accessCondition xlink:simpleLink="http://kops.ub.uni-
konstanz.de/depositlicense">http://kops.ub.uni-
konstanz.de/depositlicense</mods:accessCondition>

```

```

<mods:subject>
<mods:topic>DNA polymerases</mods:topic>
</mods:subject>
<mods:subject>
<mods:topic>NMR spectroscopy</mods:topic>
</mods:subject>
<mods:subject>
<mods:topic>non-natural amino acids</mods:topic>
</mods:subject>
<mods:subject>
<mods:topic>protein engineering</mods:topic>
</mods:subject>
<mods:subject>
<mods:topic>trifluoromethionine</mods:topic>
</mods:subject>
<mods:classification authority="ddc">540</mods:classification>
<mods:titleInfo>
A Highly Active DNA Polymerase with a Fluorous Core
</mods:titleInfo>
<mods:genre>article</mods:genre>
</xmlData>
</mdWrap>
</dmdSec>
<amdSec ID="TMD_urn_nbn_de_bsz_352-opus-102908"/>
<fileSec>
<fileGrp USE="ORIGINAL">
<file ID="urn:nbn:de:bsz:352-opus-102908_1" MIMETYPE="application/pdf"
SIZE="395469" CHECKSUM="fa5ec62503cdaca2250ab9211e3086ad"
CHECKSUMTYPE="MD5" OWNERID="http://kops.ub.uni-
konstanz.de/bitstream/urn:nbn:de:bsz:352-opus-
102908/1/Holzberger_Angew_Chem_Int_Ed_2010.pdf"
GROUPID="GROUP_urn:nbn:de:bsz:352-opus-102908_1">
<FLocat LOCTYPE="URL" xlink:type="simple" xlink:href="http://kops.ub.uni-
konstanz.de/bitstream/urn:nbn:de:bsz:352-opus-
102908/1/Holzberger_Angew_Chem_Int_Ed_2010.pdf"/>
</file>
</fileGrp>
</fileSec>
<structMap>
<div/>
</structMap>
</mets>
</metadata>

```



```

</record>
</GetRecord>
</OAI-PMH>

```

## A.4 Datensatz *Edocs* HU

Beispielsatz HU (Datenformat: XMetaDissPlus):

<http://edoc.hu-berlin.de/OAI-2.0?verb=GetRecord&metadataPrefix=xMetaDissPlus&identifier=oai:HUBerlin.de:36824>

```

<OAI-PMH xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2011-07-31T14:33:40Z</responseDate>
  <request verb="GetRecord" metadataPrefix="xMetaDissPlus"
  identifier="oai:HUBerlin.de:36944">http://edoc.hu-berlin.de/OAI-2.0</request>
  <GetRecord>
    <record>
      <header>
        <identifier>oai:HUBerlin.de:36944</identifier>
        <datestamp>2010-08-27</datestamp>
        <setSpec>pub-type:article</setSpec>
        <setSpec>open_access</setSpec>
        <setSpec>ddc:400</setSpec>
        <setSpec>ddc:490</setSpec>
      </header>
      <metadata>
        <xMetaDiss:xMetaDiss xsi:schemaLocation="http://www.d-nb.de/standards/xmetadissplus/
http://www.d-nb.de/standards/xmetadiss/xmetadissplus.xsd">
          <dc:title xsi:type="ddb:titleISO639-2" lang="eng">Out-of-focus Encoding in Gur and
Kwa</dc:title>
          <dc:creator xsi:type="pc:MetaPers">
            <pc:person>
              <pc:name type="nameUsedByThePerson">
                <pc:foreName>Anne</pc:foreName>
                <pc:surName>Schwarz</pc:surName>
              </pc:name>
            </pc:person>
          </dc:creator>
          <dc:subject xsi:type="xMetaDiss:DDC-SG">400</dc:subject>
          <dc:subject xsi:type="xMetaDiss:DDC-SG">490</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">ex-situ focus</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">focus marker</dc:subject>
          <dc:subject xsi:type="xMetaDiss:noScheme">relative clause</dc:subject>

```

```

<dc:subject xsi:type="xMetaDiss:noScheme">conjunction</dc:subject>
<dc:subject xsi:type="xMetaDiss:noScheme">grammaticalization</dc:subject>
<dc:subject xsi:type="xMetaDiss:noScheme">Afrikanische Sprachen</dc:subject>
<dcterms:abstract xsi:type="ddb:contentISO639-2" lang="eng">
This paper investigates the structural properties of morphosyntactically marked focus
constructions, focussing on the often neglected non-focal sentence part in African tone
languages. Based on new empirical evidence from five Gur and Kwa languages, we claim
that these focus expressions have to be analysed as biclausal constructions even though they
do not represent clefts containing restrictive relative clauses. First, we relativize the partly
overgeneralized assumptions about structural correspondences between the out-of-focus part
and relative clauses, and second, we show that our data do in fact support the hypothesis of
a clause coordinating pattern as present in clause sequences in narration. It is argued that we
deal with a non-accidental, systematic feature and that grammaticalization may conceal such
basic narrative structures.
</dcterms:abstract>
<dc:publisher xsi:type="cc:Publisher" type="dcterms:ISO3166">
<cc:universityOrInstitution>
<cc:name>
Humboldt Universität zu Berlin, Philosophische Fakultät II
</cc:name>
<cc:place>Berlin</cc:place>
</cc:universityOrInstitution>
<cc:address>kvv</cc:address>
</dc:publisher>
<dcterms:issued xsi:type="dcterms:W3CDTF">2005</dcterms:issued>
<dc:type xsi:type="dini:PublType">article</dc:type>
<dc:type xsi:type="dcterms:DCMIType">Text</dc:type>
<dc:identifier xsi:type="urn:nbn">urn:nbn:de:kobv:11-100173543</dc:identifier>
<dcterms:medium xsi:type="dcterms:IMT">application/pdf</dcterms:medium>
<dc:language xsi:type="dcterms:ISO639-2">eng</dc:language>
<dc:rights xsi:type="ddb:noScheme"/>
<dcterms:accessRights xsi:type="ddb:access" ddb:type="dcterms:URI"
ddb:kind="domain">
http://edoc.hu-berlin.de/e\_autoren/index-oa.php#copyrights
</dcterms:accessRights>
<thesis:degree>
<thesis:level>thesis.doctoral</thesis:level>
<thesis:grantor xsi:type="cc:Corporate">
<cc:universityOrInstitution>
<cc:name>Humboldt Universität zu Berlin</cc:name>
<cc:place>Berlin</cc:place>
</cc:universityOrInstitution>
</thesis:grantor>

```

```

</thesis:degree>
<ddb:fileNumber>1</ddb:fileNumber>
<ddb:fileProperties ddb:fileName="27GgLtqiiwXQ.pdf" ddb:fileID="file1"/>
<ddb:transfer ddb:type="dcterms:URI">
http://edoc.hu-berlin.de/oa/articles/reGh205bcnwLY/PDF/27GgLtqiiwXQ.pdf
</ddb:transfer>
<ddb:identifier ddb:type="URL">
http://edoc.hu-berlin.de/oa/articles/reGh205bcnwLY/PDF/27GgLtqiiwXQ.pdf
</ddb:identifier>
<ddb:rights ddb:kind="free"/>
</xMetaDiss:xMetaDiss>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

## A.5 Datensatz ZORA

### Beispielsatz ZORA (Datenformat: METS)

<http://www.zora.uzh.ch/cgi/oai2?verb=GetRecord&metadataPrefix=mets&identifier=oai:www.zora.uzh.ch:36700>

```

<mets:mets OBJID="eprint_36700" LABEL="Eprints Item"
xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd http://www.loc.gov/mods/v3
http://www.loc.gov/standards/mods/v3/mods-3-3.xsd" >
<mets:metsHdr CREATEDATE="2011-07-31T13:03:45Z" >
<mets:agent ROLE="CUSTODIAN" TYPE="ORGANIZATION" >
<mets:name>Zurich Open Repository and Archive</mets:name>
</mets:agent>
</mets:metsHdr>
<mets:dmdSec ID="DMD_eprint_36700_mods" >
<mets:mdWrap MDTYPE="MODS" >
<mets:xmlData>
<mods:titleInfo>
<mods:title>How harmful are survey translations? A test with Schwartz's human values
instrument</mods:title>
</mods:titleInfo>
<mods:name type="personal" >
<mods:namePart type="given" >E</mods:namePart>
<mods:namePart type="family" >Davidov</mods:namePart>
<mods:role>
<mods:roleTerm type="text" >author</mods:roleTerm>

```

```

</mods:role>
</mods:name>
<mods:name type="personal" >
<mods:namePart type="given" >A</mods:namePart>
<mods:namePart type="family" >de Beuckelaer</mods:namePart>
<mods:role>
<mods:roleTerm type="text" >author</mods:roleTerm>
</mods:role>
</mods:name>
<mods:abstract>One major challenge in international survey research is to ensure the
equivalence of translated survey instruments across different cultures. In this study, we
examine empirically the extent to which equivalence of survey instruments to measure
human values can be established across cultures sharing the same language as opposed to
cultures having a different language. We expect cultures using the same language to exhibit
higher levels of equivalence. Our examination made use of a short (i.e., a 21-item) survey
instrument to measure Schwartz's human values based on data from the second and the third
rounds of the European Social Survey (ESS). The empirical results support our
expectations.</mods:abstract>
<mods:classification authority="lcc" >Institute of Sociology</mods:classification>
<mods:originInfo>
<mods:dateIssued encoding="iso8061" >2010-09-29</mods:dateIssued>
</mods:originInfo>
<mods:originInfo>
<mods:publisher>Oxford University Press</mods:publisher>
</mods:originInfo>
<mods:genre>Journal Article</mods:genre>
</mets:xmlData>
</mets:mdWrap>
</mets:dmdSec>
<mets:amdSec ID="TMD_eprint_36700" >
<mets:rightsMD ID="rights_eprint_36700_mods" >
<mets:mdWrap MDTYPE="MODS" >
<mets:xmlData>
<mods:useAndReproduction>
<p>
<strong>Deposit:</strong>
</p>
<p>In self-archiving this collection of files and associated bibliographic metadata, I grant
ZORA the right to store them and to make them available online. I understand that the
ZORA editors may, based on the copyright and embargo situation as specified in the
previous screens, transiently or permanently restrict access to the full text files (PDF).</p>
<p>Clicking on the deposit button indicates your agreement to these terms.</p>
</mods:useAndReproduction>

```

```

</mets:xmlData>
</mets:mdWrap>
</mets:rightsMD>
</mets:amdSec>
<mets:fileSec>
<mets:fileGrp USE="reference" >
<mets:file ID="eprint_36700_67118_1" SIZE="1178070"
OWNERID="http://www.zora.uzh.ch/36700/4/Davidov_deBeuckelaer_IJPOR.pdf"
MIMETYPE="application/pdf" >
<mets:FLocat LOCTYPE="URL" xlink:type="simple"
xlink:href="http://www.zora.uzh.ch/36700/4/Davidov_deBeuckelaer_IJPOR.pdf"
></mets:FLocat>
</mets:file>
</mets:fileGrp>
</mets:fileSec>
<mets:structMap>
<mets:div DMDID="DMD_eprint_36700_mods" ADMID="TMD_eprint_36700" >
<mets:fptr FILEID="eprint_36700_document_67118_1" ></mets:fptr>
</mets:div>
</mets:structMap>
</mets:mets>

```

## A.6 Datensatz *CrossRef*

[http://www.crossref.org/guestquery?queryType=doi&restype=unixref&doi=10.1002/2Fanie.200905978&doi\\_search=Search](http://www.crossref.org/guestquery?queryType=doi&restype=unixref&doi=10.1002/2Fanie.200905978&doi_search=Search)

```

<?xml version="1.0" encoding="UTF-8"?>
<doi_records>
  <doi_record key="key" owner="10.1002" timestamp="2011-01-27 06:53:59">
    <crossref>
      <journal>
        <journal_metadata language="en">
          <full_title>Angewandte Chemie International Edition</full_title>
          <abbrev_title>Angewandte Chemie International Edition</abbrev_title>
          <issn>14337851</issn>
        </journal_metadata>
        <journal_issue>
          <publication_date>
            <month>02</month>
            <day>08</day>
            <year>2010</year>
          </publication_date>

```

```

<journal_volume>
  <volume>49</volume>
</journal_volume>
<issue>7</issue>
</journal_issue>
<journal_article publication_type="full_text">
  <titles>
    <title>A Highly Active DNA Polymerase with a Fluorous Core</title>
  </titles>
  <contributors>
    <person_name contributor_role="author" sequence="first">
      <given_name>Bastian</given_name>
      <surname>Holzberger</surname>
    </person_name>
    <person_name contributor_role="author" sequence="additional">
      <given_name>Marina</given_name>
      <surname>Rubini</surname>
    </person_name>
    <person_name contributor_role="author" sequence="additional">
      <given_name>Heiko?M.</given_name>
      <surname>Möller</surname>
    </person_name>
    <person_name contributor_role="author" sequence="additional">
      <given_name>Andreas</given_name>
      <surname>Marx</surname>
    </person_name>
  </contributors>
  <publication_date media_type="online">
    <month>01</month>
    <day>13</day>
    <year>2010</year>
  </publication_date>
  <pages>
    <first_page>1324</first_page>
    <last_page>1327</last_page>
  </pages>
  <doi_data>
    <doi>10.1002/anie.200905978</doi>
    <resource>http://doi.wiley.com/10.1002/anie.200905978</resource>
  </doi_data>
</journal_article>
</journal>
</crossref>

```

```
</doi_record>
</doi_records>
```

## A.7 Datensatz *PND*

<http://d-nb.info/gnd/14012926X/about/rdf>

```
<rdf:RDF
  xmlns:geopol="http://aims.fao.org/aos/geopolitical.owl#"
  xmlns:gnd="http://d-nb.info/gnd/"
  xmlns:rdaRelGr2="http://metadataregistry.org/uri/schema/RDARelationshipsGR2/"
  xmlns:sg="http://d-nb.info/ddc-sg/"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdaGr2="http://RDVocab.info/ElementsGr2/"
  xmlns:dnb="http://d-nb.info/"
  xmlns:ddc="http://d-nb.info/ddc/class/"
  xmlns:rdaFrbr="http://RDVocab.info/uri/schema/FRBReentitiesRDA/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:relationship="http://purl.org/vocab/relationship/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <rdf:Description rdf:about="http://d-nb.info/gnd/14012926X">
    <owl:sameAs rdf:resource="http://viaf.org/viaf/103430558"/>
    <rdaGr2:identifierForThePerson>(DE-588)14012926X
      </rdaGr2:identifierForThePerson>
    <rdaGr2:professionOrOccupation rdf:resource="http://d-nb.info/gnd/4009836-9"/>
    <rdaGr2:biographicalInformation xml:lang="de">
      Diss. Fachbereich Chemie, TU Darmstadt</rdaGr2:biographicalInformation>
    <rdaGr2:placeOfBirth xml:lang="de">Köthen</rdaGr2:placeOfBirth>
    <rdaGr2:dateOfBirth>1981</rdaGr2:dateOfBirth>
    <gnd:preferredNameForThePerson>Marx, Andreas</gnd:preferredNameForThePerson>
    <gnd:preferredNameForThePerson rdf:parseType="Resource">
      <gnd:foreName>Andreas</gnd:foreName>
      <gnd:surname>Marx</gnd:surname>
      <gnd:usedRules>RAK-WB</gnd:usedRules>
    </gnd:preferredNameForThePerson>
  </rdf:Description>
</rdf:RDF>
```

## A.8 Datensatz *LCSH*

<http://id.loc.gov/authorities/sh85037007.rdf>

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
>
  <rdf:Description rdf:about="http://id.loc.gov/authorities/sh85149860#concept">
    <skos:prefLabel xml:lang="en">Zinc enzymes</skos:prefLabel>
  </rdf:Description>
  <rdf:Description rdf:about="http://id.loc.gov/authorities/sh92004219#concept">
    <skos:prefLabel xml:lang="en">Reverse transcriptase</skos:prefLabel>
  </rdf:Description>
  <rdf:Description rdf:about="http://id.loc.gov/authorities/sh89002539#concept">
    <skos:prefLabel xml:lang="en">Polymerase chain reaction</skos:prefLabel>
  </rdf:Description>
  <rdf:Description rdf:about="http://id.loc.gov/authorities/sh85037007#concept">
    <skos:altLabel xml:lang="en">DNA nucleotidyltransferases</skos:altLabel>
    <skos:altLabel xml:lang="en">Polymerases, DNA</skos:altLabel>
    <skos:altLabel xml:lang="en">Deoxyribonucleate nucleotidyltransferases
      </skos:altLabel>
    <skos:altLabel xml:lang="en">DNA-directed DNA polymerases
      </skos:altLabel>
    <skos:altLabel xml:lang="en">Nucleotidyltransferases, Deoxyribonucleate
      </skos:altLabel>
    <skos:altLabel xml:lang="en">DNA-dependent DNA polymerases</skos:altLabel>
    <skos:prefLabel xml:lang="en">DNA polymerases</skos:prefLabel>
    <skos:narrower rdf:resource="http://id.loc.gov/authorities/sh2007009104#concept"/>
    <skos:narrower rdf:resource="http://id.loc.gov/authorities/sh97007324#concept"/>
    <skos:narrower rdf:resource="http://id.loc.gov/authorities/sh92004219#concept"/>
    <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
      1986-02-11T00:00:00-04:00</dcterms:created>
    <skos:inScheme rdf:resource="http://id.loc.gov/authorities#conceptScheme"/>
    <skos:inScheme rdf:resource="http://id.loc.gov/authorities#topicalTerms"/>
    <dcterms:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
      1989-06-01T09:43:22-04:00</dcterms:modified>
    <skos:broader rdf:resource="http://id.loc.gov/authorities/sh85149860#concept"/>
    <skos:broader rdf:resource="http://id.loc.gov/authorities/sh85136913#concept"/>
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:closeMatch rdf:resource=
```



```

    "http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb12265191n"/>
    <skos:related rdf:resource="http://id.loc.gov/authorities/sh89002539#concept"/>
    <owl:sameAs rdf:resource="info:lc/authorities/sh85037007"/>
</rdf:Description>
<rdf:Description rdf:about="http://id.loc.gov/authorities/sh2007009104#concept">
    <skos:prefLabel xml:lang="en">VDJ recombinases</skos:prefLabel>
</rdf:Description>
<rdf:Description rdf:about="http://id.loc.gov/authorities/sh97007324#concept">
    <skos:prefLabel xml:lang="en">Telomerase</skos:prefLabel>
</rdf:Description>
<rdf:Description rdf:about="http://id.loc.gov/authorities/sh85136913#concept">
    <skos:prefLabel xml:lang="en">Transferases</skos:prefLabel>
</rdf:Description>
</rdf:RDF>

```

## A.9 Datensatz *SWD*

<http://d-nb.info/gnd/4150345-4/about/rdf>

```

<rdf:RDF
  xmlns:geopol="http://aims.fao.org/aos/geopolitical.owl#"
  xmlns:rdaRelGr2="http://metadataregistry.org/uri/schema/RDARelationshipsGR2/"
  xmlns:gnd="http://d-nb.info/gnd/"
  xmlns:sg="http://d-nb.info/ddc-sg/"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdaGr2="http://RDVocab.info/ElementsGr2/"
  xmlns:dnb="http://d-nb.info/"
  xmlns:ddc="http://d-nb.info/ddc/class/"
  xmlns:rdaFrbr="http://RDVocab.info/uri/schema/FRBREntitiesRDA/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:relationship="http://purl.org/vocab/relationship/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
<skos:Concept rdf:about="http://d-nb.info/gnd/4150345-4">
  <skos:altLabel xml:lang="de">EC 2.7.7.7</skos:altLabel>
  <skos:closeMatch rdf:resource=
    "http://stitch.cs.vu.nl/vocabularies/rameau/ark:/12148/cb12265191n"/>
  <skos:broader rdf:resource="http://d-nb.info/gnd/4175227-2"/>
  <dnb:hasCoordinatedConcept-of>
    <dnb:CoordinatedConcept>
      <dnb:coordination-of rdf:resource="http://d-nb.info/ddc-sg/570"/>
    </dnb:CoordinatedConcept>
  </dnb:hasCoordinatedConcept-of>
</skos:Concept>

```

```

    <dnb:coordination-of rdf:resource="http://d-nb.info/gnd/4150345-4"/>
    <dnb:det3 rdf:resource="http://d-nb.info/ddc/class/572.786"/>
  </dnb:CoordinatedConcept>
</dnb:hasCoordinatedConcept-of>
<skos:closeMatch rdf:resource="http://id.loc.gov/authorities/sh85037007#concept"/>
<skos:prefLabel xml:lang="de">DNS-abhängige-DNS-Polymerasen</skos:prefLabel>
<skos:altLabel xml:lang="de">DNA-Polymerasen</skos:altLabel>
<dcterms:identifier>(DE-588)041503457</dcterms:identifier>
<skos:altLabel xml:lang="de">DNS-Polymerasen</skos:altLabel>
<dnb:hasCoordinatedConcept-of>
  <dnb:CoordinatedConcept>
    <dnb:coordination-of rdf:resource="http://d-nb.info/ddc-sg/570"/>
    <dnb:coordination-of rdf:resource="http://d-nb.info/gnd/4150345-4"/>
    <dnb:det1 rdf:resource="http://d-nb.info/ddc/class/572.8645"/>
  </dnb:CoordinatedConcept>
</dnb:hasCoordinatedConcept-of>
<dcterms:identifier>(DE-588c)4150345-4</dcterms:identifier>
</skos:Concept>
</rdf:RDF>

```

## A.10 Datensatz *GKD*

<http://d-nb.info/gnd/5256482-4/about/rdf>

```

<rdf:RDF
  xmlns:geopol="http://aims.fao.org/aos/geopolitical.owl#"
  xmlns:gnd="http://d-nb.info/gnd/"
  xmlns:rdaRelGr2="http://metadataregistry.org/uri/schema/RDARelationshipsGR2/"
  xmlns:sg="http://d-nb.info/ddc-sg/"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdaGr2="http://RDVocab.info/ElementsGr2/"
  xmlns:dnb="http://d-nb.info/"
  xmlns:ddc="http://d-nb.info/ddc/class/"
  xmlns:rdaFrbr="http://RDVocab.info/uri/schema/FRBReentitiesRDA/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:relationship="http://purl.org/vocab/relationship/"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
>
  <rdf:Description rdf:about="http://d-nb.info/gnd/5256482-4">
    <gnd:countryCodeForTheCorporateBody>XA-DE</gnd:countryCodeForTheCorporateBody>
    <rdaRelGr2:hierarchicalSuperior rdf:resource="http://d-nb.info/gnd/2003070-8"/>
    <rdaGr2:identifierForTheCorporateBody>(DE-588b)5256482-4

```

```

</rdaGr2:identifizierForTheCorporateBody>
<rdaGr2:identifizierForTheCorporateBody>(DE-588)952478501
</rdaGr2:identifizierForTheCorporateBody>
<gnd:variantNameForTheCorporateBody>Fakultät für Chemie &lt;Konstanz&gt;
</gnd:variantNameForTheCorporateBody>
<gnd:preferredNameForTheCorporateBody>
Universität &lt;Konstanz&gt; / Fakultät für Chemie
</gnd:preferredNameForTheCorporateBody>
</rdf:Description>
</rdf:RDF>

```